

Accessing the public genome sequence data

UCSC's Genome Browser ("Golden Path")
<http://genome.ucsc.edu>

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

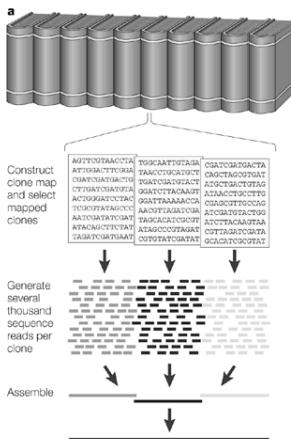
NCBI's Map Viewer
<http://www.ncbi.nlm.nih.gov/mapview/>

Types of data integrated in genome browsers

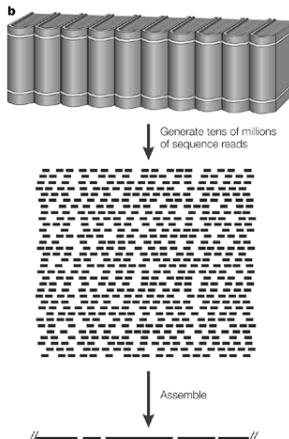
- Same starting material for all genome browsers: genomic sequence
- Annotations calculated independently by each genome browser
 - Genes
 - RefSeq mRNAs (non-redundant)
 - GenBank mRNAs (redundant)
 - ESTs
 - Gene predictions
 - SNPs
 - Homologous sequences from other organisms
 - STSs

Overview of genome sequencing strategies

Clone-by-clone shotgun sequencing



Whole-genome shotgun sequencing



Nature Reviews | Genetics
Green ED. Strategies for the systematic sequencing of complex genomes.
Nat Rev Genet. 2001; 2:573-83.

Genome Sequence Assemblies

- Complex algorithms needed to incorporate all sequence data
- Assemblies updated periodically as new sequence becomes available
 - Mouse and human genomes assembled by NCBI
 - 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-test.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/hg19/ GRCh37	Build 37.1	GRCh37
Mouse	Yes	July 2007/mm9/Build 37	Build 37.1	Build 37
Dog	Yes	May 2005 /canFam 2.0	Build 2.1/ CanFam 2.0	CanFam 2.0
Zebrafish	NO	Dec 2008/danRer6/ Zv8	Zv7/build 3.1	Zv8
Rhesus	Yes	Jan 2006/rheMac2/ v. 1.0, Mmul_051212	Build 1.1/ v.1.0, Mmul_051 212	Mmul_1

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

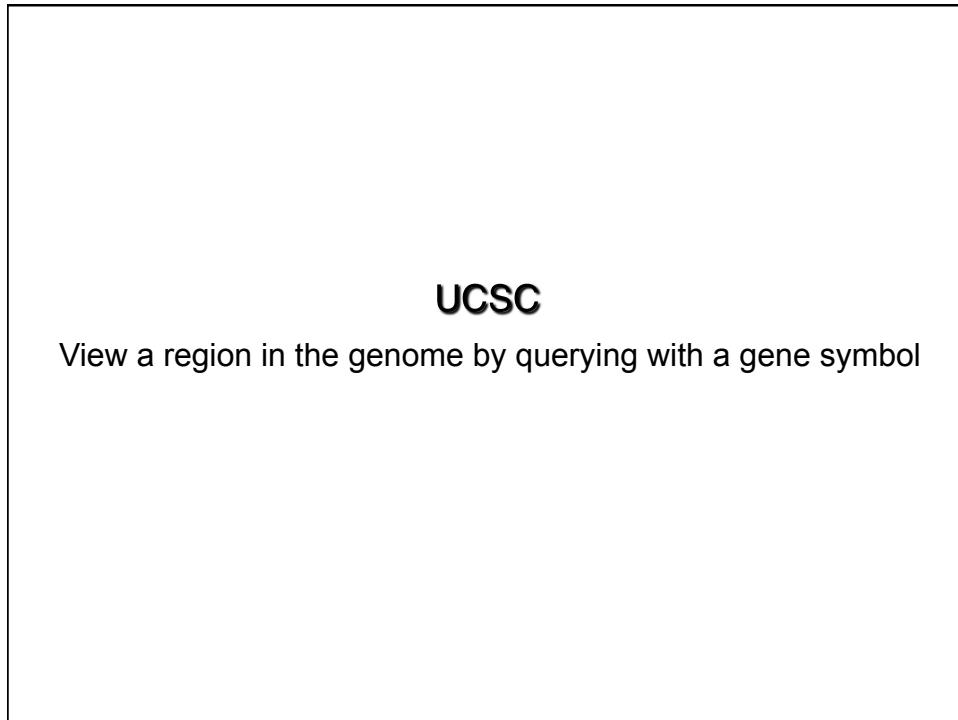
<http://www.ncbi.nlm.nih.gov/RefSeq/key.html>

Beta actin mRNA RefSeq

LOCUS NN_001101 1852 bp mRNA linear PRI 27-DEC-2009
 DEFINITION Homo sapiens actin, beta (ACTB), mRNA.
 ACCESSION NM_001101
 VERSION NN_001101.3 GI:168480144
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Barchontoglires; Primates; Haplorrhini; Catarrhini; Hominoidea; Homo.
 REFERENCE 1
 AUTHORS Yamaguchi,H., Shiraiishi,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
 J. Cell. Physiol. 222 (3), 748-755 (2009)
 PUBMED 19475567

 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.
 Summary: This gene encodes one of six different actin proteins. Actins are highly conserved proteins that are involved in cell motility, stability and integrity. The actin is a major constituent of the contractile apparatus and one of the two nonmuscle cytoskeletal actins. [provided by RefSeq].
 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.
 CDS
 85..1212
 /gene="ACTB"
 /gene_synonym="PS1TP5BP1"
 /molecule="beta cytoskeletal actin; PS1TP5-binding protein 1; actin, cytoplasmic 1"
 /codon_start=1
 /product="beta actin"
 /protein_id="NP_001102.1"
 /db_xref="Gi:4501885"
 /db_xref="GeneID:1102"
 /db_xref="HGNc:1327"
 /db_xref="HPRD:00032"
 /db_xref="MIM:102630"
 /translation="MDDDIALAVVNDNSGKCKAGFAGGDAPRFFPSIVGRPRHQGVN
 VGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 VLLTEADLNPKANREKMQIIMPETPPAMYYVAIQLVLSLYASCNTTGIVMDGCGGT
 HTPVLYSGVALPHALRLDLLAGRQDYLMLKLITERGVSFTTAEREIVRDIKEKLCY
 VALDFEQEMATAAASSSLKESYELPDPGCVITIGNERFRCPPEALFLQPSPFLGMESCCIH
 TTFNSIMKCDVWIKRDLVANTVLSGGTTMYPGCIADMQKEITLAPSTMHKIKIAPE
 RKYSWIGGSILASLSTPQQMWISKQEYDESGPSIVHRKCP"

 ORIGIN
 1 accgcggaaa ccgcgtccgc cccggggaaa cggggcgccg cttttggaa tcggccggcc
 61 ggcaacaccc ggcggggaaa caccatgggt gatataatgg cggggatgtt cgccggaaaac
 121 ggctccggaa ttttggggaa cggggcgccg cggccggatg ccccccggcc cgtttccccc



The screenshot shows the UCSC Genome Bioinformatics website at <http://genome.ucsc.edu/>. A black arrow points to the "Genomes" link in the top navigation bar, which is highlighted with a red box. The main content area displays the "About the UCSC Genome Bioinformatics Site" page. The sidebar on the left lists various tools: Genome Browser, ENCODE, Blat, Table Browser, Gene Sorter, In Silico PCR, Genome Graphs, Galaxy, VisiGene, Proteome Browser, Utilities, Downloads, Release Log, Custom Tracks, and Archaeal Genomes.

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 a portal to the ENCODE project.

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.

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](#).

25 Jan. 2010 - New Custom Track type: Binary Alignment/Map (BAM)

We are pleased to announce the availability of another new track type for Custom Tracks: the Binary Alignment/Map (BAM).

BAM is the compressed binary version of the Sequence Alignment/Map (SAM) format, a compact and indexable representation of nucleotide sequence alignments. Many next-generation sequencing and analysis tools work with SAM/BAM. For custom track display, the main advantage of indexed BAM over PSL and other human-readable alignment formats is that only the portions of the files needed to display a particular region are transferred to UCSC. This makes it possible to display alignments from files that are so large that the connection to UCSC would time out when attempting to upload the whole file to UCSC. Both the BAM file and its associated index file remain on your web-accessible server (http or ftp), not on the UCSC server. UCSC temporarily caches the accessed portions of the files to speed up interactive display.

13 Jan. 2010 - Lifespan of custom tracks within sessions: Due to the popularity of UCSC custom tracks and sessions, we are running out of disk space for storing custom tracks accessed within sessions. Although sessions

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 genome assembly position or search term image width

Mammal Human Mar. 2006 ADAM2 800 submit

GRCh37 Mar. 2006 Click here to re-set user interface settings to their defaults.
 Mar. 2006 May 2004 July 2003 add custom tracks and display clear position

About the Human Mar. 2006 (hg18) assembly (sequences)

The March 2006 human reference sequence (NCBI Build 36.1) was produced by the International Human Genome Sequencing Consortium.

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST marker, or a cytological band, 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
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	Displays region between STS markers RH18061 and RH80175 or chromosome bands 15q11 to 15q13. This syntax may also be used for other range queries, such as between uniquely-determined ESTs, mRNAs

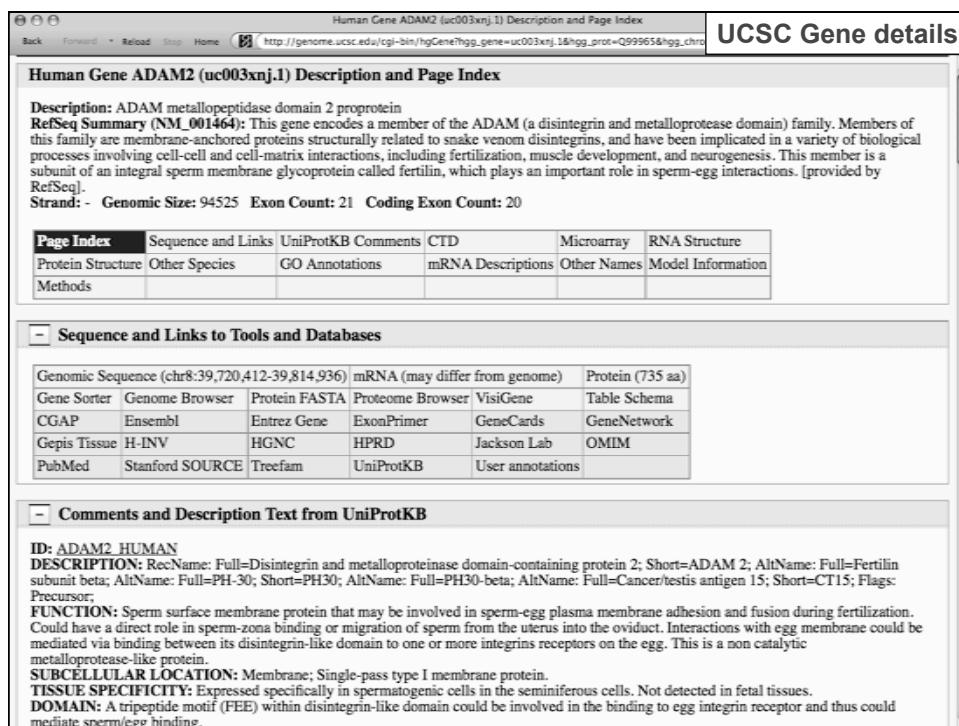
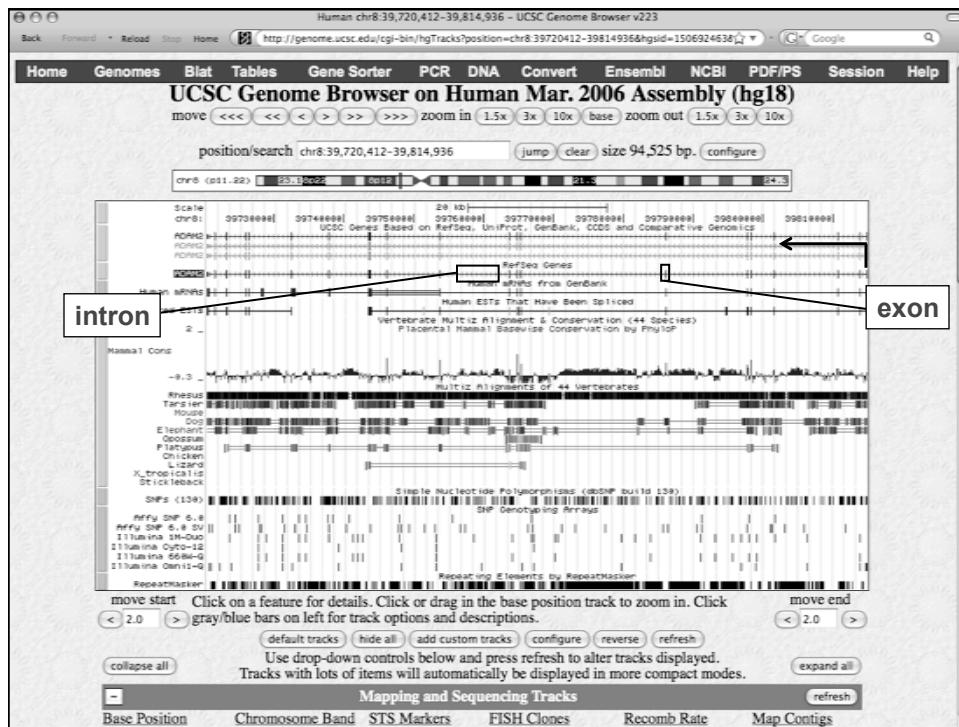
Homo sapiens
 (Graphic courtesy of CBSE)

UCSC Genes

ADAM2 (uc003xn1..1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein
 ADAM2 (uc003xnk..1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein
 ADAM2 (uc003xnk..1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein
 ADAM28 (uc0101ca..1) at chr8:24240012-24268671 - ADAM metallopeptidase domain 28 isoform 1
 ADAM29 (uc0101rr..1) at chr4:176133252-176135714 - ADAM metallopeptidase domain 29 preproprotein
 ADAM2 (uc010ard..1) at chr14:70059047-70061377 - ADAM metallopeptidase domain 20 preproprotein
 ADAM28 (uc003xdy..1) at chr8:24207525-24268671 - ADAM metallopeptidase domain 28 isoform 1
 ADAM28 (uc003xdy..1) at chr8:24207525-24268671 - ADAM metallopeptidase domain 28 isoform 3
 ADAM22 (uc003ujp..1) at chr7:87402007-87649275 - ADAM metallopeptidase domain 22 isoform 4
 ADAM22 (uc003ujp..1) at chr7:87401638-87664385 - ADAM metallopeptidase domain 22 isoform 3
 ADAM22 (uc003ujn..1) at chr7:87401638-87664385 - ADAM metallopeptidase domain 22 isoform 1
 ADAM22 (uc003ujm..1) at chr7:87401638-87664385 - ADAM metallopeptidase domain 22 isoform 2
 ADAM22 (uc003ujl..1) at chr7:87401638-87649364 - ADAM metallopeptidase domain 22 isoform 5
 ADAM22 (uc003ujk..1) at chr7:87401638-87649364 - ADAM metallopeptidase domain 22 isoform 4
 ADAM22 (uc003ujj..1) at chr7:87401638-87600698 - ADAM metallopeptidase domain 22 isoform 5
 ADAM22 (uc003ujl..1) at chr7:87401638-87600698 - ADAM metallopeptidase domain 22 isoform 5
 ADAM29 (uc003id..1) at chr4:176088712-176135906 - ADAM metallopeptidase domain 29 preprotein
 ADAM29 (uc003id..1) at chr4:176076134-176135906 - ADAM metallopeptidase domain 29 preprotein
 ADAM29 (uc003id..1) at chr4:176076134-176135906 - ADAM metallopeptidase domain 29 preprotein
 ADAM23 (uc002vq..1) at chr2:207016612-207196924 - ADAM metallopeptidase domain 23 preprotein
 ADAM20 (uc001xmd..1) at chr14:70058831-70071485 - ADAM metallopeptidase domain 20 preprotein
 ADAM21 (uc001xmd..1) at chr14:695993970-69596375 - ADAM metallopeptidase domain 21 preprotein
 YWHAZ (uc002xma..1) at chr20:42947758-42970575 - tyrosine 3-monooxygenase/trypophan
 YWHAZ (uc002xmt..1) at chr20:42947758-42970575 - tyrosine 3-monooxygenase/trypophan
 YWHAZ (uc002zfs..1) at chr17:1194593-1250267 - tyrosine 3/trypophan 5-monooxygenase
 YWHAZ (uc003alz..1) at chr22:30670479-30683590 - tyrosine 3-monooxygenase/trypophan
 YWHAQ (uc002gqx..1) at chr2:9641557-9688196 - tyrosine 3/trypophan 5-monooxygenase
 YWHAZ (uc003yjw..1) at chr8:102000090-10203475 - tyrosine 3/trypophan 5-monooxygenase
 YWHAZ (uc003yjw..1) at chr8:102000090-102033447 - tyrosine 3/trypophan 5-monooxygenase
 YWHAZ (uc003yjw..1) at chr8:102000090-102032853 - tyrosine 3/trypophan 5-monooxygenase
 YWHAZ (uc010mr..1) at chr8:102000090-102034287 - tyrosine 3/trypophan 5-monooxygenase
 ADAM21P (uc010arb..1) at chr14:69782223-69784271 - ADAM21-like protein.

RefSeq Genes

ADAM2 at chr8:39720412-39814936 (NM_001464) ADAM metallopeptidase domain 2 proprotein
 ADAM20 at chr14:70058831-70071485 (NM_003814) ADAM metallopeptidase domain 20 preprotein
 ADAM21 at chr14:695993970-69596375 (NM_003813) ADAM metallopeptidase domain 21 preprotein
 ADAM21P1 at chr14:69782223-69784271 (NM_003951)
 ADAM22 at chr7:87401638-87649364 (NM_004194) ADAM metallopeptidase domain 22 isoform 4
 ADAM22 at chr7:87401638-87664385 (NM_021723) ADAM metallopeptidase domain 22 isoform 1
 ADAM22 at chr7:87401638-87664385 (NM_021722) ADAM metallopeptidase domain 22 isoform 2



Affymetrix All Exon Microarrays

Region	Fold Energy	Bases	Energy/Base	Display As
5' UTR	-15.00	75	-0.200	Picture PostScript Text
3' UTR	-72.02	359	-0.201	Picture PostScript Text

The RNAfold program from the Vienna RNA Package is used to perform the secondary structure predictions and folding calculations. The estimated folding energy is in kcal/mol. The more negative the energy, the more secondary structure the RNA is likely to have.

Protein Domain and Structure Information

InterPro Domains: Graphical view of domain structure

- IPR006586 - ADAM_Cys-rich
- IPR001762 - Blood-coag_inhib_Disintegrin
- IPR018358 - Disintegrin_CS
- IPR013032 - EGF-like_reg_CS
- IPR013111 - EGF_extracell
- IPR001590 - Peptidase_M12B
- IPR002870 - Peptidase_M12B_N

Pfam Domains:

- PF01421 - Reprolysin (M12B) family zinc metalloprotease
- PF01562 - Reprolysin family propeptide
- PF08516 - ADAM cysteine-rich
- PF00200 - Disintegrin
- PF07974 - EGF-like domain

SCOP Domains:

RefSeq Gene ADAM2

UCSC RefSeq Gene details

RefSeq: NM_001464_3 Status: Reviewed
 Description: Homo sapiens ADAM metallopeptidase domain 2 (ADAM2), mRNA.
 CCDS: CCDS34884.1
 CDS: 3 complete
 OMIM: 601533
 Entrez Gene: 2515
 PubMed on Gene: ADAM2
 PubMed on Product: ADAM metallopeptidase domain 2 proprotein
 GeneCards: ADAM2
 AceView: ADAM2
 Stanford SOURCE: NM_001464
 CDS FASTA alignment from multiple alignment: NM_001464

Summary of ADAM2

This gene encodes a member of the ADAM (a disintegrin and metalloprotease domain) family. Members of this family are membrane-anchored proteins structurally related to snake venom disintegrins, and have been implicated in a variety of biological processes involving cell-cell and cell-matrix interactions, including fertilization, muscle development, and neurogenesis. This member is a subunit of an integral sperm membrane glycoprotein called fertilin, which plays an important role in sperm-egg interactions. [provided by RefSeq].

mRNA/Genomic Alignments

SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
2642	100.0%	8	-	39720412	39814936	NM_001464	1	2642	2657

.....

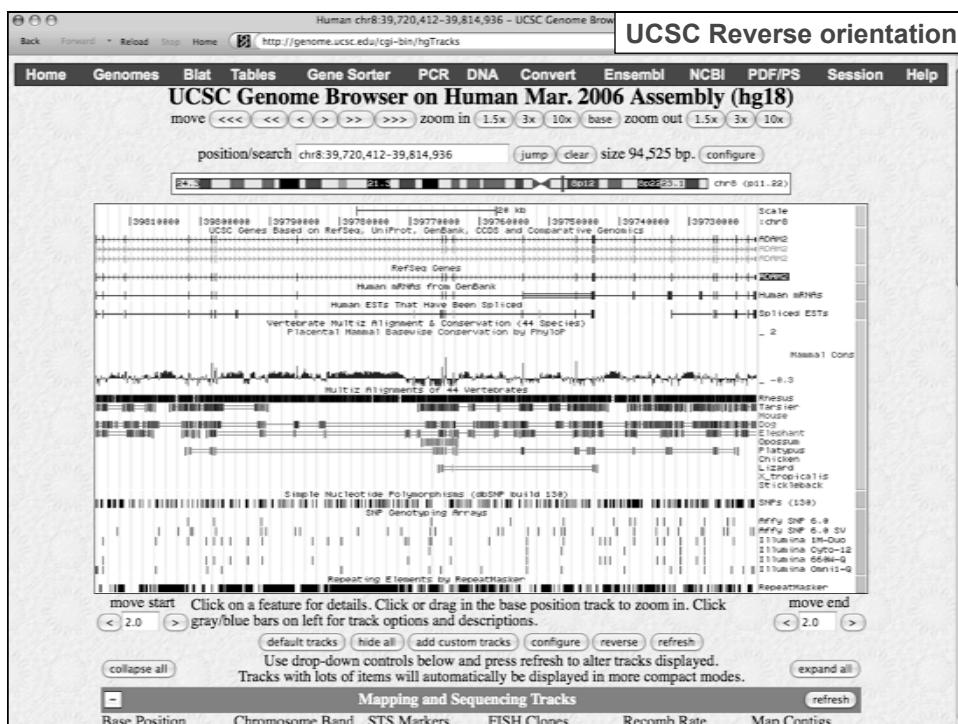
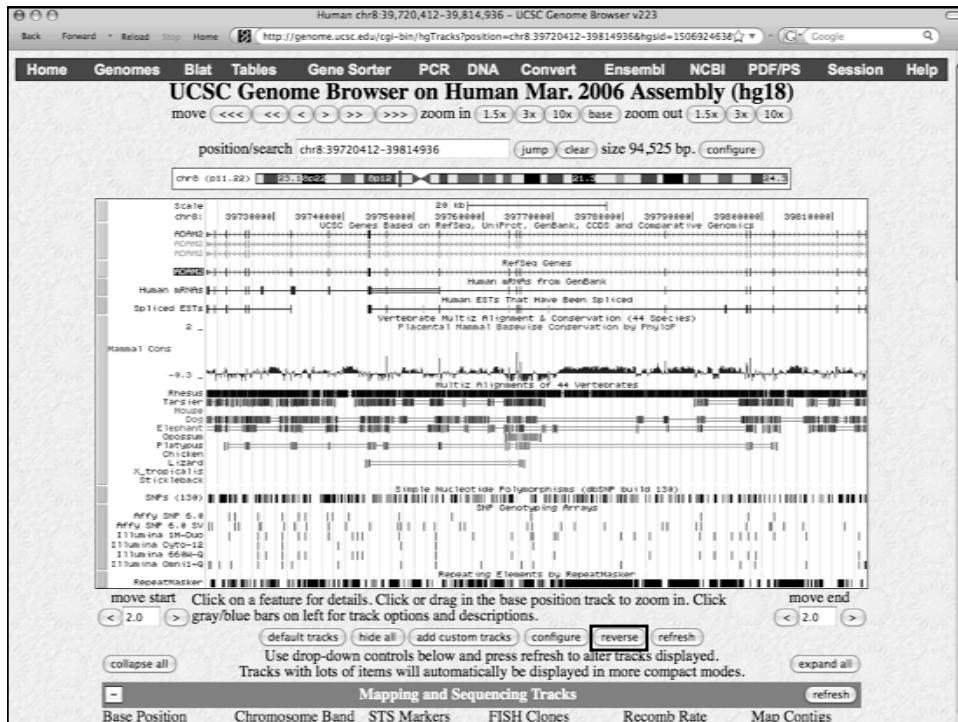
Links to sequence:

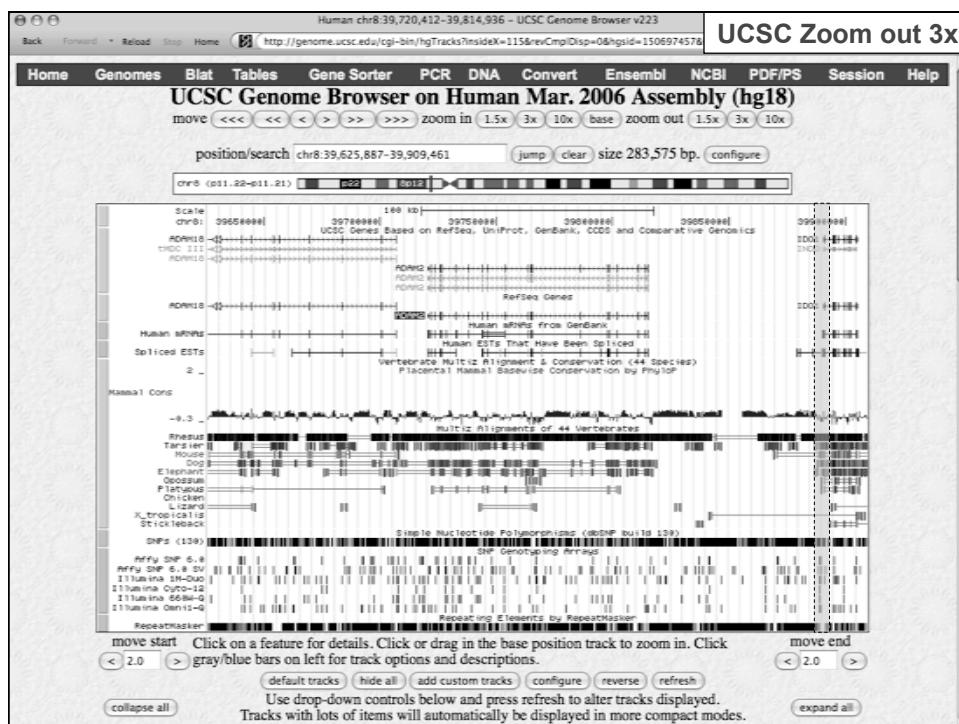
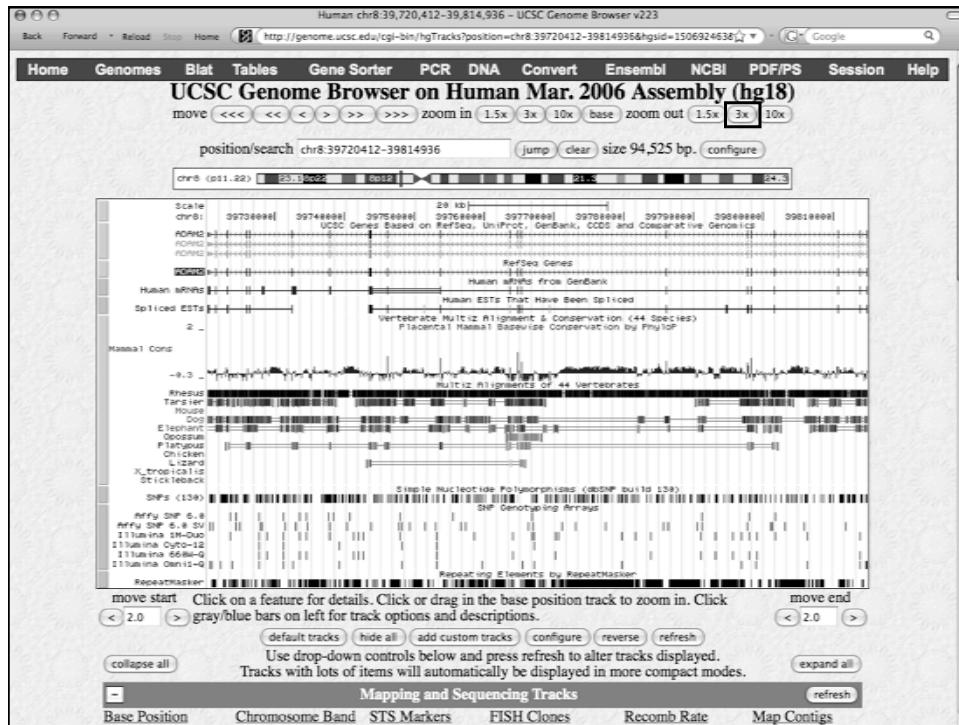
- Predicted Protein
- mRNA Sequence m different from the genomic sequence.
- Genomic Sequence from assembly

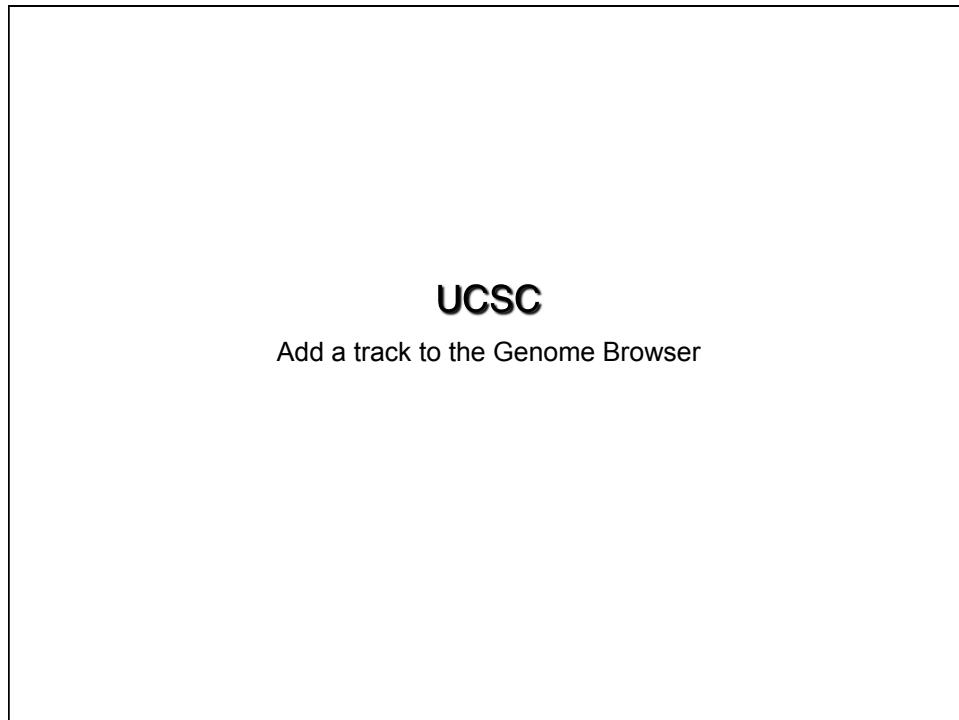
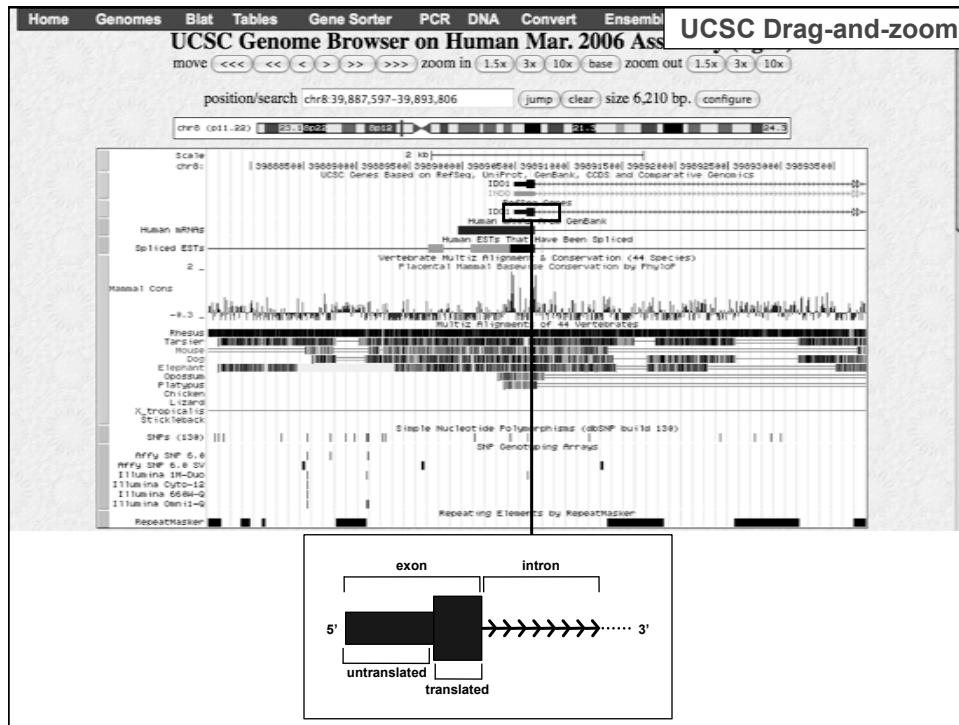
click

The screenshot shows the UCSC RefSeq Gene details page for the gene ADAM2. The main title is "Genomic Sequence Near Gene". Below it, a section titled "Get Genomic Sequence Near Gene" contains a note about using the Table Browser for multiple features. A "Sequence Retrieval Region Options" section includes checkboxes for Promoter/Upstream by 1000 bases (checked), 5' UTR Exons, CDS Exons, 3' UTR Exons, Introns, Downstream by 1000 bases, One FASTA record per gene (checked), and One FASTA record per region (exon, intron, etc.) with Split UTR and CDS parts of an exon into separate FA. A note below states that features close to the chromosome edge may extend beyond the edge. The "Sequence Formatting Options" section includes radio buttons for Exons in upper case, CDS in upper case, All upper case, All lower case, and Mask repeats (radio button checked). A "submit" button is present. To the right, a large text box displays the genomic sequence: "1000 nt upstream of ADAM2" followed by the DNA sequence itself, which is a long string of nucleotides starting with >hg18_refGene_NM_001464_range=chr8:39814937-39815936.

The image shows the UCSC logo at the top center. Below it, the text "Navigating around the Genome Browser" is displayed in a serif font.







Human chr8:39,720,222-39,723,392 – UCSC Genome Browser v223

UCSC track selection

Phenotype and Disease Associations

Genes and Gene Prediction Tracks

UCSC Genes	Old UCSC Genes	Alt Events	Gencode Genes	CCDS	RefSeq Genes
pack	hide	hide	hide	hide	pack
Other RefSeq	MGC Genes	ORFeome Clones	TransMap...	Vega Genes	Ensembl Genes
hide	hide	hide	hide	hide	hide
AccView Genes	SIB Genes	N-SCAN	CONTRAST	SGP Genes	Geneid Genes
hide	hide	hide	hide	hide	hide
Genscan Genes	Exoniphy	Augustus	RNA Genes	ACEScan	EvoFold
hide	hide	hide	hide	hide	hide
sno/miRNA	Pos Sel Genes				
hide	hide				

mRNA and EST Tracks

Human mRNAs	Spliced ESTs	Human ESTs	Other mRNAs	Other ESTs	H-Inv
dense	dense	hide	hide	hide	hide
UniGene	Gene Bounds	SIB Alt-Splicing	Poly(A)	CGAP SAGE	
hide	hide	hide	hide	hide	

Expression

Broad Histone	CpG Islands	E10/JCVI NAS	Epontine TSS	FirstEF	GIS ChIP-PET
hide	hide	hide	hide	hide	hide
HAIB Methyl-seq	HAIB Methyl27	HAIB TFBS	NHGRIBi-Pro	NHGRINRE	Open Chromatin
hide	hide	hide	hide	[No data-chr8]	hide
RegAnno	SUNY RBP	SwitchGear TSS	TFBS Conserved	TS miRNA sites	UW DNasel HS
hide	hide	hide	hide	hide	hide
VisTa	Yale TFBS	7X Reg Potential	FOX2 CLIP-seq	LI/UCSD TAF1...	NKI Nuc Laminin
hide	hide	hide	hide	hide	hide
Nucleosome Occupancy	hIP...				
occupancy	full				

Regulation

Broad Histone	CpG Islands	E10/JCVI NAS	Epontine TSS	FirstEF	GIS ChIP-PET
hide	hide	hide	hide	hide	hide
HAIB Methyl-seq	HAIB Methyl27	HAIB TFBS	NHGRIBi-Pro	NHGRINRE	Open Chromatin
hide	hide	hide	hide	[No data-chr8]	hide
RegAnno	SUNY RBP	SwitchGear TSS	TFBS Conserved	TS miRNA sites	UW DNasel HS
hide	hide	hide	hide	hide	hide
VisTa	Yale TFBS	7X Reg Potential	FOX2 CLIP-seq	LI/UCSD TAF1...	NKI Nuc Laminin
hide	hide	hide	hide	hide	hide
Nucleosome Occupancy	hIP...				
occupancy	full				

Comparative Genomics



Yale TFBS Track Settings

Maximum display mode: full

Select views (help):

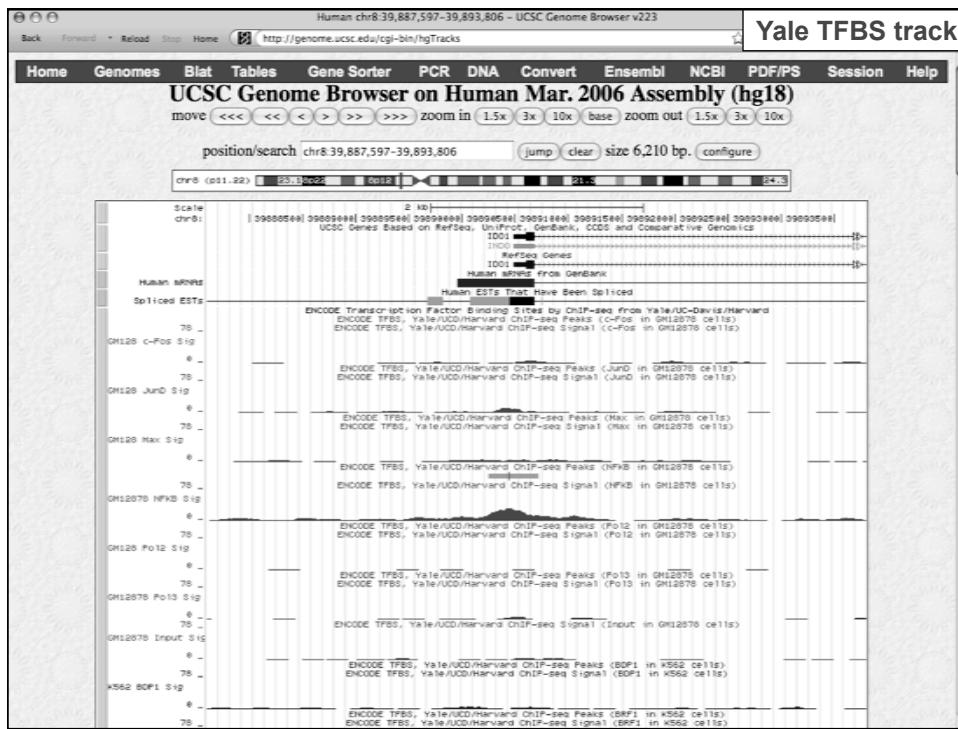
Peaks

Peaks Configuration

Minimum Q-Value (-log 10):	0	(0 to 300)
Minimum P-Value (-log 10):	0	(0 to 300)
Minimum Signal value:	0	(0 to 18241)

Select subtracks by cell line and factor:

All		Cell Line	GM12878	K562	HeLa-S3	HepG2	HCT116	HEK293(b)	NB4	NT2-D1	(NTera2)	HeLa-S3	K562	K562	K562
Factor	+	-	+	-	+	-	+	-	+	-	+	+	+	+	+
AP-2alpha	+/-														
AP-2gamma	+/-														
BDP1	+/-														
BRF1	+/-														
c-Fos	+/-														
c-Jun	+/-														
c-Myc	+/-														
E2F1 (HA-E2F1)	+/-														
E2F4	+/-														
E2F6	+/-														
GATA-1	+/-														
GATA-2	+/-														
JunD	+/-														



UCSC

Change the color of items in a track

Human chr8 39,887,597-39,893,806 – UCSC Genome Browser v223

Back Forward Reload Stop Home http://genome.ucsc.edu/cgi-bin/hgTracks Google

Lamprey Net Medaka Chain Medaka Net Stickleback Chain Stickleback Net Fugu Chain
 hide ▾ hide ▾ hide ▾ hide ▾ hide ▾ hide ▾
 Fugu Net Tetraodon Chain Tetraodon Net Tetraodon Ecores Zebrafish Chain Zebrafish Net
 hide ▾ hide ▾ hide ▾ hide ▾ hide ▾ hide ▾
 X. tropicalis Chain X. tropicalis Net Zebra finch Chain Zebra finch Net Lizard Chain Lizard Net
 hide ▾ hide ▾ hide ▾ hide ▾ hide ▾ hide ▾
 Platypus Chain Platypus Net Chimp Chain/Net Orangutan Chain/Net Rhesus Chain/Net Guinea Pig Chain
 hide ▾ hide ▾ hide ▾ hide ▾ hide ▾ hide ▾
 Guinea Pig Net Rat Chain Rat Net Marmoset Chain/Net Mouse Chain Mouse Net
 hide ▾ hide ▾ hide ▾ hide ▾ hide ▾ hide ▾
 Dog Chain/Net Cat Chain/Net Horse Chain/Net Cow Chain/Net Opossum Chain/Net Chicken Chain/Net
 hide ▾ hide ▾ hide ▾ hide ▾ hide ▾ hide ▾

Variation and Repeats refresh

C SNPs (130) SNPs (129) SNPs (128) SNPs (126) SNP Arrays
 CNV dense hide ▾ hide ▾ hide ▾ hide ▾ dense ▾
 hide ▾

HGDP Allele Freq HGDP Smoothed HGDP Hetzygosity HGDP iHS HGDP XP-EHH HapMap SNPs
 FST hide ▾ hide ▾ hide ▾ hide ▾ hide ▾
 hide ▾

HapMap LD Phased Tajima's D SNPs Tajima's D DGV Struct Var HGSV Discordant Segmental Dups
 hide ▾ hide ▾ hide ▾ hide ▾ hide ▾ hide ▾
 Structural Var Exapted Repeats RepeatMasker RepMask 3.2.7 Interrupted Rpts Intr Rpts 3.2.7
 hide ▾ hide ▾ dense ▾ hide ▾ hide ▾ hide ▾
 Simple Repeats Microsatellite Self Chain Genome Variants
 hide ▾ hide ▾ hide ▾ hide ▾

Pilot ENCODE Regions and Genes refresh

Pilot ENCODE Transcription refresh

Pilot ENCODE Chromatin Immunoprecipitation refresh

Pilot ENCODE Chromatin Structure refresh

Pilot ENCODE Comparative Genomics and Variation refresh

(refresh)

SNPs (130) Track Settings UCSC SNP Track details

Simple Nucleotide Polymorphisms (dbSNP build 130)

Display mode: pack

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

On details page, show function and coding differences relative to:

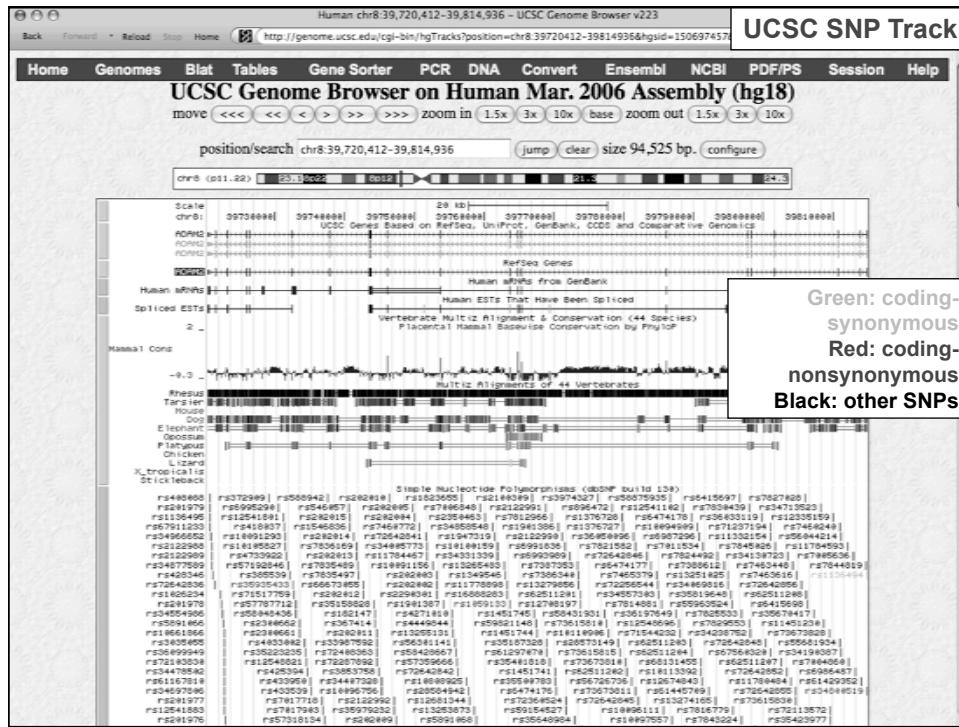
<input checked="" type="checkbox"/> UCSC Genes	<input type="checkbox"/> Old UCSC Genes	<input type="checkbox"/> Gencode Manual	<input type="checkbox"/> Gencode Auto
<input type="checkbox"/> Gencode PolyA	<input type="checkbox"/> CCDS	<input type="checkbox"/> RefSeq Genes	<input type="checkbox"/> Other RefSeq
<input type="checkbox"/> Vega Protein Genes	<input type="checkbox"/> Vega Pseudogenes	<input type="checkbox"/> Ensembl Genes	<input type="checkbox"/> AceView Genes
<input type="checkbox"/> SIB Genes	<input type="checkbox"/> N-SCAN PASA-EST	<input type="checkbox"/> N-SCAN	<input type="checkbox"/> SGP Genes
<input type="checkbox"/> Geneid Genes	<input type="checkbox"/> Genscan Genes	<input type="checkbox"/> Exoniphy	<input type="checkbox"/> Augustus Hints
<input type="checkbox"/> Augustus De Novo	<input type="checkbox"/> Augustus Ab Initio	<input type="checkbox"/> ACEScan	

Minimum Average Heterozygosity: 0
 Maximum Weight: 3 SNPs with higher weights are less reliable

SNP Feature for Color Specification: Function

The selected feature above has the following values below. For each value, a selection of colors is available. If a SNP has more than one of these properties, resulting in more than one color, then the stronger color will override the weaker color. In order from strongest to weakest, the colors are red, green, blue, gray, black.

Unknown <input type="button" value="black +"/>	Locus <input type="button" value="black +"/>	Coding - Synonymous <input type="button" value="green +"/>	Coding - Non-Synonymous <input type="button" value="red +"/>
Untranslated <input type="button" value="black +"/>	Intron <input type="button" value="black +"/>	Splice Site <input type="button" value="black +"/>	



UCSC

Find a chicken homolog of a human protein

NCBI Entrez Protein

Protein – ADAM metallopeptidase domain 2 proprotein [Homo sapiens]

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Protein for Go Clear

Format: GenPept FASTA Graphics More Formats▼

Downloaded Save Links▼

NCBI Reference Sequence: NP_001455.3

ADAM metallopeptidase domain 2 proprotein [Homo sapiens]

```
>gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
MGRVLFLLSCLGGLRMDNSFDLPLWQTYKEKIRSLIKBOIESQASXKIVIEBCKPVTVLNMQKMLPENF
RVTSYSGTGMKFLDQPQNPCHYCHQYLBGPXPSVVMVSTCIGLRLQVLFENVSYGIEPLESSVGFEEVII
YQVKHHKKADVSLYNEKDLEISRLSFKLQSVVPQOFAXYIEMHVIVEKQLYNHMGMSDTZVVAQKVNQLIG
LTNAIFVSFVNPI111SLSLWIDENK1ATTEGEANELLHTYLWKTTSVYLRLPHDVAPLLVYREKSNVYGA
TFPGKMKCDAKYAGGVVPTPTISLESLAVILAQQLSLSMNTTDDINKECQCSGAVWMMPEFAIHFSGVKI
FSNCSCFEDPAHPF1SREKQSCQDQMQDQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ
FVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ
SGQOCICDTGKEVEYGPSECYLSLNLSTIDWGNCCGSDCYTQCEADNLQCGKLIKCYVGKFLLQJFRA
TIIYANISGHLCIAVEFASDWDASDQSKMVKDGTSOGSNXVCRMQCBGSQSYLGVDCTTDCKNDNRGVCNKXK
KHCCHCSASYLPPDCSVQSDLMPGGSIDSGNFPFPVVAIPARLPERRYIENLYBSKPKMRWPPFLFIPFFIIFC
VLIAIMVKVNFQRKKWRTEDYSSDQESESEPEKG
```

Change Region Shown

Analyze This Sequence

- Run BLAST
- Identify Conserved Domains

Articles about the ADAM2 gene

- Mapping, sequence, and expression analysis of the human fertlin beta gene [Genomics, 1997]
- Role of the integrin-associated protein CD9 in binding beta [Proc Natl Acad Sci U S A, 1999]
- Mediation of sperm-egg fusion: evidence that mouse egg alpha6beta1 integrin [Cell Biol, 1999]

» See all...

Identical Proteins for NP_001455.3

- unnamed protein product [Homo sa] [CBH30599]
- ADAM metallopeptidase domain 2 [EAW63273]
- RecName: Full=Disintegrin and [O09965]

» See all...

RefSeq mRNA

See reference mRNA sequence for the ADAM2 gene (NM_001464.3).

More about the ADAM2 gene

This gene encodes a member of the ADAM (a disintegrin and metalloprotease domain) family. Members of this family are membrane-anchored

UCSC BLAT search

Chicken BLAT Search

Home Genomes Tables PCR Session FAQ Help

Chicken BLAT Search

BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

```
>gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
MGRVLFLLSCLGGLRMDNSFDLPLWQTYKEKIRSLIKBOIESQASXKIVIEBCKPVTVLNMQKMLPENF
RVTSYSGTGMKFLDQPQNPCHYCHQYLBGPXPSVVMVSTCIGLRLQVLFENVSYGIEPLESSVGFEEVII
YQVKHHKKADVSLYNEKDLEISRLSFKLQSVVPQOFAXYIEMHVIVEKQLYNHMGMSDTZVVAQKVNQLIG
LTNAIFVSFVNPI111SLSLWIDENK1ATTEGEANELLHTYLWKTTSVYLRLPHDVAPLLVYREKSNVYGA
TFPGKMKCDAKYAGGVVPTPTISLESLAVILAQQLSLSMNTTDDINKECQCSGAVWMMPEFAIHFSGVKI
FSNCSCFEDPAHPF1SREKQSCQDQMQDQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ
FVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ
SGQOCICDTGKEVEYGPSECYLSLNLSTIDWGNCCGSDCYTQCEADNLQCGKLIKCYVGKFLLQJFRA
TIIYANISGHLCIAVEFASDWDASDQSKMVKDGTSOGSNXVCRMQCBGSQSYLGVDCTTDCKNDNRGVCNKXK
KHCCHCSASYLPPDCSVQSDLMPGGSIDSGNFPFPVVAIPARLPERRYIENLYBSKPKMRWPPFLFIPFFIIFC
VLIAIMVKVNFQRKKWRTEDYSSDQESESEPEKG
```

submit I'm feeling lucky clear

In a query sequence to find its location in the genome. Multiple sequences may be searched if separated by lines starting with > followed by the sequence name.

File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence:

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.

For locating PCR primers, use In-Silico PCR for best results instead of BLAT.

About BLAT

Chicken BLAT Results

UCSC BLAT search

Home Genomes Tables PCR Session FAQ Help

Chicken BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CER0	STRAND	START	END	SPAN
Browser details	NP_001455.3	44	539	600	735	71.6%	Un	++	635370	635555	186
browser details	NP_001455.3	12	301	304	735	100.0%	1	++	67659709	67659720	12
browser details	NP_001455.3	12	437	440	735	100.0%	1	++	67660117	67660128	12
browser details	NP_001455.3	12	385	390	735	83.4%	1	++	67659961	67659978	18

UCSC Genome Browser on Chicken Feb. 2004 Assembly (galGal2)

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

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

Chicken BLAT Results

UCSC BLAT search

Home Genomes Tables PCR Session FAQ Help

Chicken BLAT Results

BLAT Search Results

ACTIONS QU Alignment of NP_001455.3 and chrUn:635370-635555

Browser details	NP	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.	
Browser details	NP		
Browser details	NP		
Browser details	NP		

NP_001455.3

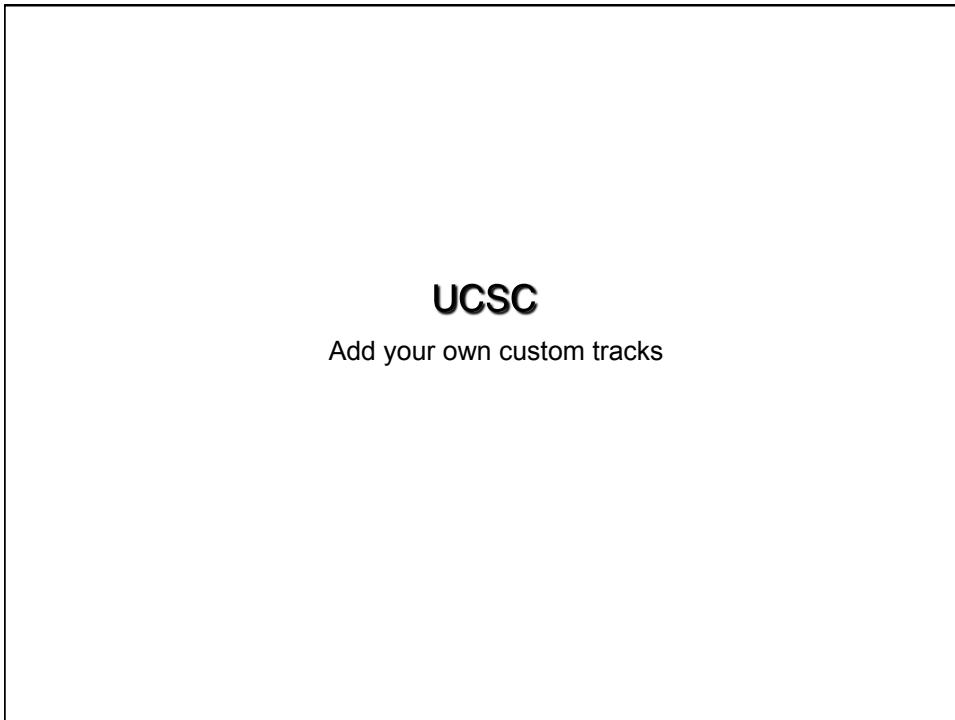
```
mrvrlfllesg lgglrmdsnf dslpvqitvp ekirsiikeq iesqasykiv iegkpytvnl 60
mzkfnlphnf rvysygtgi mpkldqdfqn fchyggyieg ypksvvmwst ctgjrgvlf 120
envsygiepl essvgfehvi ygvkhkkadv slynekdies rdlsfkqlgsy eppqdfakyi 180
emhviwekl ynhmgsdttv vaqkvflig ltnaifvsn itiilsslel widenkiatt 240
gsselihft rwtksyivl rphdaffil yreksnyvgv tifqgmdan yaggvvilhpr 300
tisqslavl lagsqslav lqddinckndv sccqkqslav qzqslav fscnclslav 360
hfiskepl lqddpdrdpf fqsgavavma klasgssedde gtedocslig qzsdidatav 420
fkqeqnnaeq pocencifms kormcrpsfe ecdfiyeyng ssasacpenhy vgtghpcgn 480
qwicidgvcm sgdkqctdtf qkewefgpfz cyslnqnciidsd gytqceadL 540
qCGKLICK-v qkflqjipra "IXYAnishH L" iavefasd hadsqkmwiK DGTsCGsnK/ 600
crnqrvcsws yigydcttdk cndrrgvcnnk khchcasayl ppdsqvndl wpggsidegr 660
fppvaiparl perryienly hskpmrwppf lfipffiiife vliaimvkvn fgrkwrkwd 720
yssdeqpepe sepkq
```

Chicken.chrUn :

```
AACTGGggcT GTGGAAACT CATCTGCaca TA:ccaaacc gagttccctt caccaatta 635429
aaagggtCCA TCATCTATGC Tcaagtgcga gaaCATCTGT G:ggtgttctt tgatgtatg 635489
catgcaccc ttccccccat ccggcagaca tcctctctg gttAGGATG GCACGaaaTG CGGTccccga 635549
AAGGTA
```

Side by Side Alignment*

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



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

Human (*Homo sapiens*) Genome Browser Gateway

UCSC Custom Tracks

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 genome assembly position or search term image width
 Mammal Human Mar. 2006 chrX:151,073,054-151,383,976 800 submit

Click here to reset the browser user interface settings to their defaults.
 add custom tracks configure tracks and display clear position

Add Custom Tracks

clade genome assembly [hg18]
 Mammal Human Mar. 2006 [hg18]

Display your own data as custom annotation tracks in the browser. Data must be formatted in BED, b
 bigWig, MAF, BAM or PSL formats. To configure the display, set track and browser line attributes as d
 data in the bigBed and bigWig formats must be embedded in a track line in the box below. Publicly
 Examples are here.

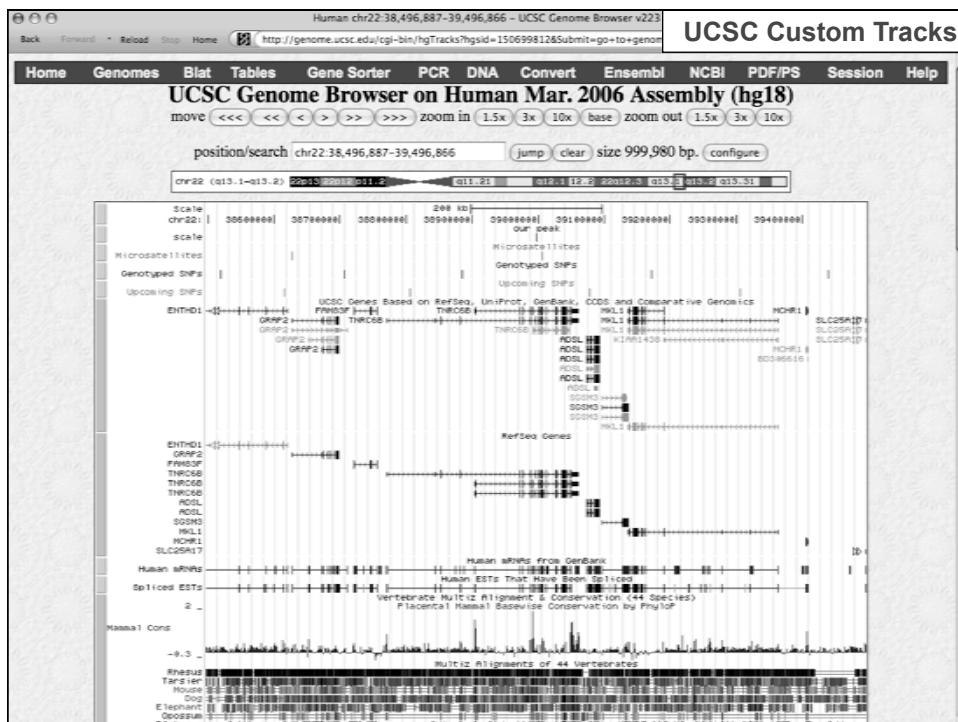
Paste URLs or data: Or upload: Browse... Submit
 chr22 39171390 39171391 ss22916
 chr22 39438769 39438770 ss1479794
 track name="Upcoming SNPs" description="Upcoming SNPs" color=0,128,192
 chr22 38615712 38615713 ss86855
 chr22 38804838 38804839 ss85533
 chr22 39077895 39077896 ss141190
 chr22 39305065 39305066 ss137027

Manage Custom Tracks

genome: Human assembly: Mar. 2006 [hg18]

Name	Description	Type	Doc	Items	Pos	delete
scale	our peak	bed		1	chr22:	<input type="checkbox"/>
Microsatellites	Microsatellites	bed		2	chr22:	<input type="checkbox"/>
Genotyped SNPs	Genotyped SNPs	bed		5	chr22:	<input type="checkbox"/>
Upcoming SNPs	Upcoming SNPs	bed		4	chr22:	<input type="checkbox"/>

add custom tracks go to genome browser go to table browser
 check all / clear all (+ -)



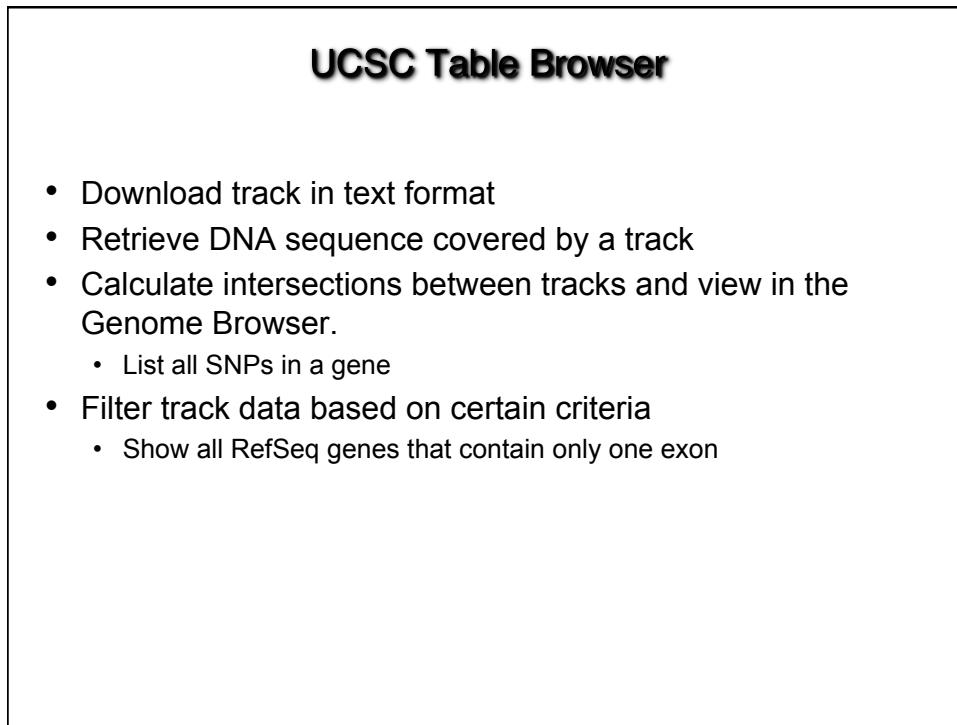
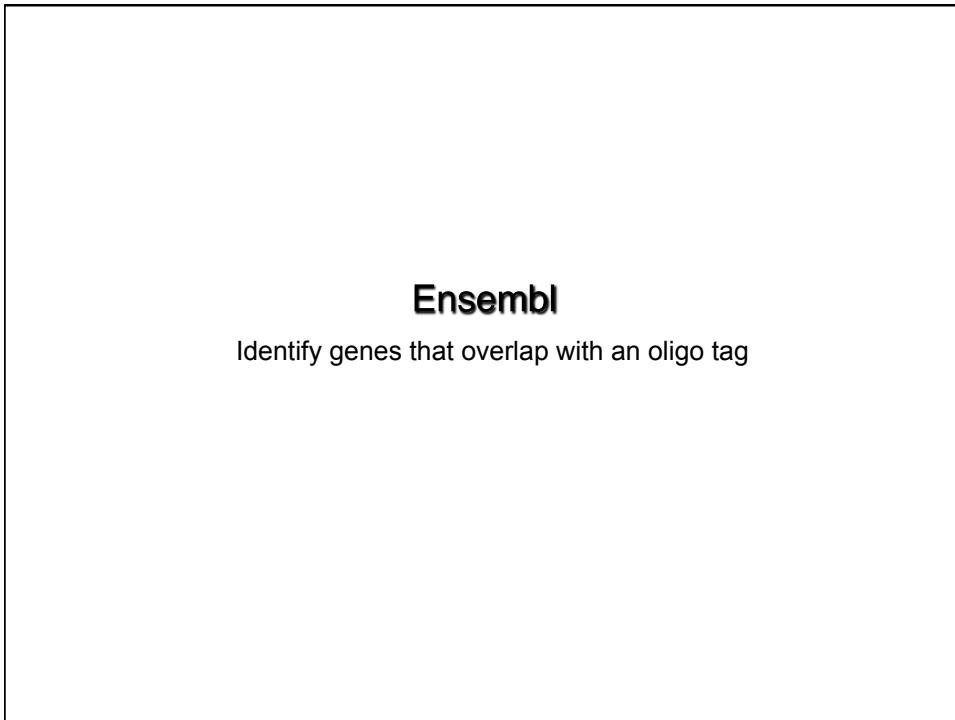


Table Browser															UCSC Table Browser: RefSeq genes that contain only one exon							
Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks and view in the Genome Browser. For help in using this application see Using the Table Browser for a detailed User's Guide for general information and sample queries, and the OpenHelix Table Browser tutorial for features and usage. For more complex queries, you may want to use Galaxy or our public MySQL server.															Filter on Fields from hg18.refGene							
clade: <input type="button" value="Mammal"/> genome: <input type="button" value="Human"/> assembly: <input type="button" value="Mar. 2004"/> group: <input checked="" type="radio" value="Genes and Gene Prediction Tracks"/> <input type="radio" value="track"/> table: <input checked="" type="radio" value="refGene"/> <input type="radio" value="describe table schema"/> region: <input checked="" type="radio" value="genome"/> <input type="radio" value="ENCODE"/> <input type="radio" value="position"/> position chr22:38496887- identifiers (names/acccessions): <input type="text" value="paste list"/> <input type="button" value="upload list"/> filter: <input type="button" value="create"/> intersection: <input type="button" value="create"/> correlation: <input type="button" value="create"/> output format: <input type="button" value="all fields from selected table"/> output file: <input type="text"/> (leave blank to) file type returned: <input checked="" type="radio" value="plain text"/> <input type="radio" value="gzip compressed"/> <input type="button" value="get output"/> <input type="button" value="summary/statistics"/> To reset all user cart settings (including custom tracks), click here															Filter on Fields from hg18.refGene bin is ignored AND 0 name does match * chrom does match * strand does match * txStart is ignored AND 0 txEnd is ignored AND 0 cdsStart is ignored AND 0 cdsEnd is ignored AND 0 exonCount is = 1 AND exonStarts does match * exonEnds does match * id is ignored AND 0 name2 does match * cdsStartStat does match AND cdsEndStat does match AND #filter: refGene.exonCount = 1 #bin name chrom strand txStart txEnd cdsStart cdsEnd exonCount exonStarts exonEnds id name2 cdsStartStat cdsEndStat exonFrames 88 NM_006511 chr1 + 15858950 15860804 15858950 15860804 1 15858950, 15860804 0 RSC1A1 cmpl cmpl <0, 178 NM_002232 chr1 - 111015832 111019178 111017226 11101955 1 111015832, 111019178, 0 KCNA3 cmpl cmpl <0, 301 NM_001821 chr1 - 239858789 239865955 239865959 239865970 1 239858789, 239865955, 0 CHML cmpl cmpl <0, 585 NM_001054848 chr1 + 59853 59853 59853 59853 1 59853, 59853 0 OR4F5 cmpl cmpl <0, NM_001052771 chr1 + 357521 358460 357521 358460 1 357521, 358460, 0 OR4F16 cmpl cmpl <0, 587 NM_001005221 chr1 + 357521 358460 357521 358460 1 357521, 358460, 0 OR4F29 cmpl cmpl <0, NM_001005224 chr1 + 357521 358460 357521 358460 1 357521, 358460, 0 OR4F3 cmpl cmpl <0, NM_001005277 chr1 - 610958 611897 610958 611897 1 610958, 611897, 0 OR4F16 cmpl cmpl <0, 589 NM_001005221 chr1 - 610958 611897 610958 611897 1 610958, 611897, 0 OR4F29 cmpl cmpl <0, NM_001005224 chr1 - 610958 611897 610958 611897 1 610958, 611897, 0 OR4F3 cmpl cmpl <0, NR_031741 chr1 - 556128 556128 556128 556128 1 556128, 556128, 0 MDR1P77 unk unk unk <0, 590 NR_024321 chr1 - 752449 752765 752765 752765 1 752449, 752765, 0 MCRN400115 unk unk unk <0, 593 NR_029639 chr1 + 1092346 1092441 1092441 1092441 1 1092346, 1092441, 0 MGR2008 unk unk unk <0,							



You've been redirected to your nearest mirror - uswest.ensembl.org

Take me back to www.ensembl.org

click

Ensembl Genome Browser

Home

Login / Register | BLAST/BLAT | BioMart | Docs & FAQs | Mirrors

Search: All species for Co

e.g. human gene BRCA2 or rat X:100000..200000 or insulin

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online. Click on a link below to go to the species' home page.

Popular genomes ([Log in to customize this list](#))

Human GRC37

Mouse NCBIM37

Zebrafish Zv8

All genomes

-- Select a species --

[View full list of all Ensembl species](#)

Other pre-build species are available in [Ensembl PreI](#).

New to Ensembl?

Did you know you can:

- Learn how to use Ensembl with our video tutorials and walk-throughs
- Add custom tracks using our new Control Panel
- Upload your own data and save it to your Ensembl account
- Search for a DNA or protein sequence using BLAST or BLAT
- Fetch only the data you want from our public database, using the Ensembl Perl API
- Download our databases via FTP in FASTA, MySQL and other formats
- Mine Ensembl with BioMart and export sequences or tables in text, html, or Excel format

Did you know...?

A preliminary assembly of the common baboon (*Papio hamadryas*) is now available on our pre-site, <http://pre.ensembl.org/Baboon>

Still got questions? [Try our FAQs](#)

What's New in Release 56 (15 September 2009)

- New species - pig (Pig)
- New species - marmoset (Marmoset)
- New rat gene set (Bat)

The screenshot shows the 'Ensembl BLAST search' interface. At the top, it says 'BlastView' and 'Enter the Query Sequence'. A sequence is pasted into the input field: >NPSS_1 AAAAATGTCGGCTGAAGAG. To the right, there's a sidebar with sections for 'setup', 'configure', 'results', and 'Ensembl BLAST search' (which is currently selected). Below the query sequence, there's a link to 'http://research.nhgri.nih.gov/teaching/custom_tracks.shtml'. The main form includes fields for 'Enter a sequence ID or accession (EMBL, UniProt, RefSeq)' and 'Or Enter an existing ticket ID:'. It also has radio buttons for 'DNA queries' (selected) and 'Peptide queries'. Under 'Select the databases to search against', 'Species' dropdown shows 'Gasterosteus_aculeatus', 'Gorilla_gorilla', and 'Homo_sapiens'. 'Database' dropdown shows 'LATESTGP' (selected) and 'PEP_ALL'. Under 'Select the Search Tool', 'BLASTN' is selected. In the 'Search sensitivity' dropdown, 'Near-exact matches (oligo)' is selected. Other options include 'Exact matches', 'Near-exact matches', 'Allow some local mismatch', 'Distant homologies', and 'No optimisation'. A note at the bottom states: 'BlastView provides an integrated platform for sequence similarity searches against Ensembl'.

The screenshot shows the results of a BLAST search. At the top, it says 'BlastView' and 'Showing top 100 alignments of 2, sorted by Raw Score'. The results are displayed in three panels: 'Alignment Locations vs. Karyotype', 'Alignment Locations vs. Query', and 'Alignment Summary'. The 'Alignment Locations vs. Karyotype' panel shows a genome browser-like view of chromosomes with black and white segments indicating alignment coverage. The 'Alignment Locations vs. Query' panel shows a bar chart of HGP coverage. The 'Alignment Summary' panel is a table with columns: Subject, Chromosome, Supercontig, Clone, Contig, Chromosome, Chromosome, Chromosome, Sort By, Start, End, Ori, Score, E-val, and Length. A large arrow points from the 'Query Start' and 'End' values (1 and 20) in the summary table back up to the 'Alignment Locations vs. Query' panel. A smaller arrow points from the 'Sort By' column in the summary table to the 'Sort By' dropdown in the summary panel. The table data is as follows:

Link	Start	End	Ori	Chromosome	Name	Start	End	Ori	Chromosome	Name	Start	End	Ori	Chromosome	Name	Start	End	Ori	Score	E-val	Length		
[A] [S] [G] [C]	1	20	+	Chr:15	57210876	57210895	+	Chr:15	57210876	57210895	+	Chr:15	57210876	57210895	+	Chr:8	72042559	72042575	+	20	0.0093	100.00	20
[A] [S] [G] [C]	1	17	-	Chr:8	72042559	72042575	+	Chr:8	72042559	72042575	+	Chr:8	72042559	72042575	+	Chr:8	72042559	72042575	+	17	0.57	100.00	17

**Ensembl Location tab:
 Region in detail**

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail**
- Comparative Genomics
- Alignments (image) (5)
- Alignments (text) (51)
- Multi-species view (47)
- Synteny (12)
- Genetic Variation
- Resequencing (2)
- Linkage Data
- Markers
- Other genome browsers
- UCSC
- NCBI

- Configure this page
- Manage your data
- Export data
- Bookmark this page

Location: 15:57,208,876-57,212,895

Chromosome 15: 57,208,876-57,212,895

Region overview Region in detail Alignments (image) *

Chromosome bands Contigs Ensembl/Havana g. ncRNA gene ncRNA pseudogene

Gene Legend

Ensembl Homo sapiens version 56.97a (GRCh37) Chromosome 15: 56,720,896 - 57,710,895

Known protein coding Novel protein coding Known RNA gene Novel RNA gene

Location: 15:57,208,876-57,212,895 Export Image

Chromosome bands Human RefSeqExon

Genes → ← Genes

BLAT/BLAST hits Contigs Human RefSeqExon Reg. Feats

Gene Legend

Known protein coding Untranslated

translated exon untranslated exon

Ensembl

Location tab

Ensembl Location tab: Region in detail

Ensembl genome browser 56: H.sapiens – Region in detail – Chromosome 15: 57,208,876-57,212,895

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
- Alignments (image) (5)
- Alignments (text) (51)
- Multi-species view (47)
- Synteny (12)
- Genetic Variation
- Resequencing (2)
- Linkage Data
- Markers
- Other genome browsers
- UCSC
- NCBI

Configure this page

Manage your data

Export data

Bookmark this page

click

Location: 15:57,208,876-57,212,895

Region overview Region in detail help Alignments (image) *

Chromosome bands Contigs Ensembl|Havana g... ncRNA gene ncRNA pseudogene Gene Legend

Ensembl Homo sapiens version 56.97a (GRCh37) Chromosome 15: 56,710,086 - 57,710,085 Known protein coding Novel protein coding Novel RNA gene

BLAT/BLAST hits Contigs Human RefSeqEN... CCDS set Ensembl|Havana g...

BLAT/BLAST hits Contigs Human RefSeqEN... Reg. Feats Gene Legend

Known protein coding Unclassified Reverse strand Known protein coding Unclassified

Ensembl Location tab: Configure page

Ensembl genome browser 56: H.sapiens – Region in detail – Chromosome 15: 57,208,876-57,212,895

Configure page Main panel Top panel Custom Data Your account Save and close

Search display:

Main panel

Active tracks

(1/4) Sequence

(0/1) Markers

(4/11) Genes

(0/2) Prediction Transcripts

(0/6) Protein alignments

(0/5) Protein features

(2/4) cDNA/mRNA alignments

(0/4) EST alignments

(0/2) RNA alignments

(0/1) Other DNA alignments

(0/26) Probe features

(0/3) Dtag features

(0/1) External data

(0/4) Simple features

(0/12) Misc. regions

(0/19) Variation features

(1/23) Functional genomics

(0/8) Multiple alignments

(0/36) BLASTZ alignments

(0/11) Translated blast alignments

(5/5) Additional decorations

(5/5) Information

Variation features

All variants (Show info)

Off (Show info)

Normal (Show info)

chip 100K Array variations (Show info)

array 500K Array variations (Show info)

Affy GenomeWideSNP_6.0 variations (Show info)

EGA variations (Show info)

ENSEMBL-Venter variations (Show info)

ENSEMBL-Watson variations (Show info)

Illumina_CytoSNP12v1 variations (Show info)

Illumina_HumanM-duoV3 variations (Show info)

Illumina_Human60W-quad variations (Show info)

NHGRI_GWAS_catalog variations (Show info)

Uniprot variations (Show info)

dbSNP variations (Show info)

B48 ASTD human SNPs (Show info)

B48 CONDOR human (Show info)

B48 DECIPHER (Show info)

B48 DGV loci (Show info)

B48 WGTP regions (Show info)

To update this configuration, select your tracks and other options in the box above and close this popup window. Your view will then be updated automatically.
 Reset configuration for Main panel to default settings.

Notes:

- Change whether a track is drawn OR how it is drawn, click on the icon by the track name and then select the way the track is to be rendered.
- On the left hand side of the name the number of tracks in a menu, and the number of tracks currently turned on from

Gene Legend

Known protein coding Unclassified Reverse strand Known protein coding Unclassified

**Ensembl Location tab:
 Region in detail with additional features**

Ensembl genome browser 56: H.sapiens - Region in detail - Chromosome 15: 57,208,876-57,212,895

Chromosome bands

Contigs

Ensembl/Havana g...

ncRNA gene

ncRNA pseudogene

Gene Legend

Location: 15 : 57208876 - 57212895 [Go]

Chromosome bands

Human RefSeq/EN... CCCS set

Ensembl/Havana g...

BLAST/BLAST hits

Contigs

All variations

Step Heats

57,208,000 57,208,500 57,210,000 57,210,500 57,211,000 57,211,500 57,212,000 57,212,500

Forward strand

Reverse strand

Known protein coding

Novel pseudogene

Novel RNA gene

TCF12-002 > Known protein coding Ensembl/Havana merge gene

TCF12-001 > Known protein coding Ensembl/Havana merge gene

TCF12-202 > Known protein coding Ensembl gene

TCF12-203 > Known protein coding Ensembl gene

AC080599.6 >

Export Image

<< click >>

Forward strand

**Ensembl Location tab:
 Region in detail after navigation**

Ensembl genome browser 56: H.sapiens - Region in detail - Chromosome 15: 57,212,896-57,216,905

Chromosome bands

Human RefSeq/EN... CCCS set

Ensembl/Havana g...

ncRNA gene

ncRNA pseudogene

Gene Legend

Location: 15 : 57212896 - 57216905 [Go]

Chromosome bands

Human RefSeq/EN... CCCS set

Ensembl/Havana g...

Contigs

All variations

Step Heats

57,213,000 57,213,500 57,214,000 57,214,500 57,215,000 57,215,500 57,216,000 57,216,500

Forward strand

Reverse strand

Known protein coding

Novel pseudogene

Novel RNA gene

TCF12-002 > Known protein coding Ensembl/Havana merge gene

TCF12-001 > Known protein coding Ensembl/Havana merge gene

TCF12-202 > Known protein coding Ensembl gene

TCF12-203 > Known protein coding Ensembl gene

AC080599.6 >

Variation: rs35615435

Gene Legend

Variation Legend

Configuring the display

You currently have 15 tracks in the main panel turned off. To change the tracks you are displaying, use the "Configure this panel" button at the bottom of the page.

Ensembl release 56 - Sept 2009 © W

Permanent link - View in archive site

Properties

bp: 57213283

status: freq

class: intron

ambiguity: R

code:

mapweight: 1

alleles: A/G

source: Illumina_CytoSNP12v1, dbSNP

type: SYNONYMOUS_CODING

Synonymous coding

About Ensembl | Contact Us | Help

Export Image

Ensembl Variation tab: Summary

Variation: rs35615435

Variation class: SNP (source dbSNP)

Synonyms: Illumina_CytoSNP12v1 rs35615435

Alleles: A/G (Ambiguity code: R)

Location: This feature maps to 1 genomic location(s). [hide locations](#)

15:57213283 (forward strand) [Jump to region in detail](#)

Variation summary [help](#) [Gene/Transcript >](#)

Validation status: Proven by frequency (Feature tested and validated by a non-computational method).

Linkage disequilibrium data: No linkage data for this SNP

Flanking Sequence: AATCTTGGACTTTCCATGTAATGTCGGGCTTCAGAACAAAATGTACCCATTATTC
 CAATTAAATTTAGAATAATGATCATTCAGTGATTTGCCCTCTGGATCTGAATGTTATC
 CAATTAAATTTAGAATAATGATCATTCAGTGATTTGCCCTCTGGATCTGAATGTTATC
 TGGATCTGGATCTGGATCTGGATCTGGATCTGGATCTGGATCTGGATCTGGATCTGG
 GAATTTTCCCGCCCTTAAATGATGGGAAACATAGGCACATACACTGGGAGCAGCA
 RTCAAGTGGGATCAGGATAGAGATGATGCTTAAACACTAAAGACTCATATTGGTGGTGGAT
 ACATTTTAGTAGAAATAATGAGAAAGCATACATTAAATATTTGATATAATGATGATATAA
 AAGATGGAAATTGACTCAAGTGTATAATGTAATGTTAAATTAAGATGACAGTT
 CCGCTGAAATCTTGGGAAACTCAAGTGTGGATCTGGATCTGGATCTGGATCTGG
 GTAAATTTCAAATGTAATGAAACTCTAGATTATTTAAATGTAATGATGATGAC
 T

(Variant highlighted)

Ensembl release 56 - Sept 2009 © [WTSI](#) / [EBI](#) [About Ensembl](#) | [Contact Us](#) | [Help](#)

Permanent link - [View in archive site](#)

Ensembl Variation tab: Context

Variation: rs35615435

Variation class: SNP (source dbSNP)

Synonyms: Illumina_CytoSNP12v1 rs35615435

Alleles: A/G (Ambiguity code: R)

Location: This feature maps to 1 genomic location(s). [hide locations](#)

15:57213283 (forward strand) [Jump to region in detail](#)

Individual genotypes [Context](#) [help](#) [Phenotype Data >](#)

Contigs: All variations

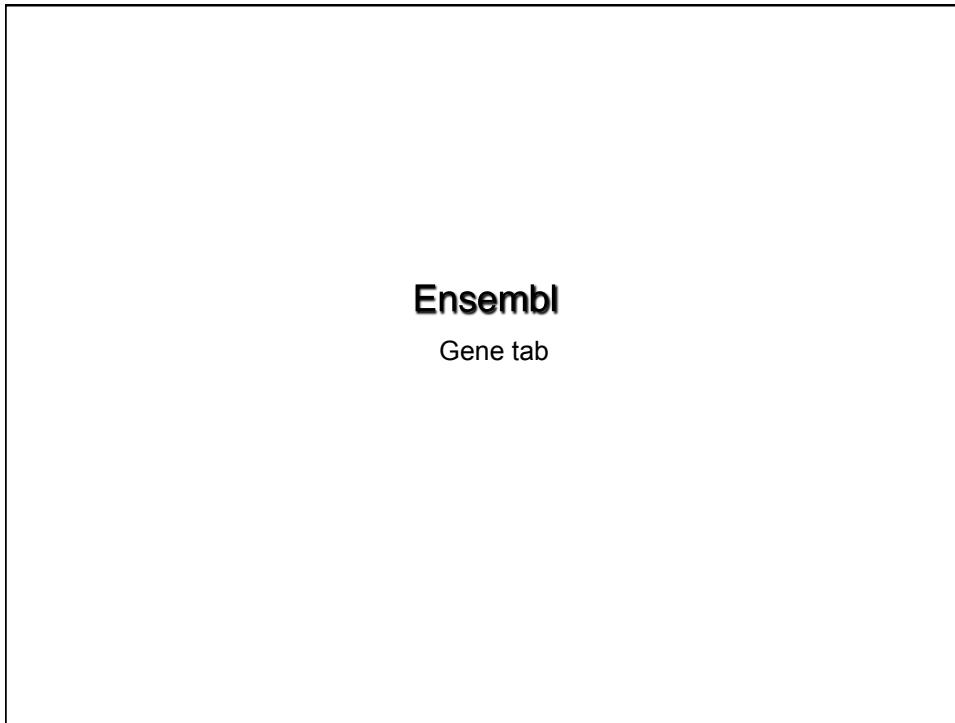
Upstream [Forward](#)

30.00 Kb Accession # 57.20 Mb 57.21 Mb 57.22 Mb 57.23 Mb

Variation legend:
 ■ Intergenic
 ■ Synonymous coding
 ■ Reverse strand
 ■ Intronic

Ensembl release 56 - Sept 2009 © [WTSI](#) / [EBI](#) [About Ensembl](#) | [Contact Us](#) | [Help](#)

Permanent link - [View in archive site](#)



**Ensembl Location tab:
 Region in detail**

This screenshot shows the Ensembl Location tab for the "Region in detail" view. The main panel displays a genomic track for Chromosome 15, specifically the region between 57,212,896 and 57,216,915. The track includes various genomic features like exons, introns, and gene models, with labels such as AC090518.2, AC090517.2, and TCF12. A black arrow points to the "TCF12-001" entry in the track list, which is highlighted with a yellow background. The left sidebar contains a navigation menu with links like "Back", "Forward", "Reload", "Home", and "Configure this page". The bottom of the page has a footer with copyright information.

Ensembl Gene tab: Gene summary

Ensembl genome browser 56: H.sapiens - Gene summary - Gene: TCF12 (ENSG00000140262)

Gene-based displays Gene: TCF12 (ENSG00000140262)

Transcription Factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) Source: UniProtKB/Swiss-Prot

Location Chromosome 15: 57,210,823-57,582,051, forward strand.

Transcripts There are 5 transcripts in this gene; [hide transcripts](#).

Name	Transcript ID	Protein ID	Description
TCF12-001	ENST00000267811	ENSP00000267811	protein_coding
TCF12-002	ENST00000333725	ENSP00000331057	protein_coding
TCF12-201	ENST00000342822	ENSP00000342459	protein_coding
TCF12-202	ENST00000438423	ENSP00000388940	protein_coding
TCF12-203	ENST00000452095	ENSP00000396881	protein_coding

Transcript and Gene level displays

In Ensembl a gene is made up of one or more transcripts. We provide displays at two levels:

- Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
- Gene views which provide displays for data associated at the gene level such as orthologues and paralogues, regulatory regions and splice variants.

This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.

Gene summary help Splice variants *

Name TCF12 (HGNC (curated))
 Synonyms bHLHb20, HEB, HsT17266, HTF4 [[\[To view all Ensembl genes linked to the name click here\]](#)]
 CCDS This gene is a member of the Human CCDS set: CCDS10159, CCDS10160, CCDS42042
 Gene type Known protein coding

Prediction MethodGene containing both Ensembl genebuild transcripts and [Havana](#) manual curation, see article.
 Alternative genes This Known protein coding entry corresponds to the following database identifiers:
 Havana Gene: OTTHUMG00000132047 [[view all locations](#)]

Transcripts

Ensembl(Havana 9) 57.25 Mb 57.30 Mb 57.35 Mb 57.40 Mb 57.45 Mb 57.50 Mb 57.55 Mb

TCF12-002 > Known protein coding Ensembl/Havana merge gene
 TCF12-001 > Known protein coding Ensembl/Havana merge gene
 TCF12-201 > Known protein coding Ensembl gene
 TCF12-202 > Known protein coding Ensembl gene
 TCF12-203 > Known protein coding Ensembl gene

Ensembl Gene tab: Orthologues

Ensembl genome browser 56: H.sapiens - Orthologues - Gene: TCF12 (ENSG00000140262)

Gene-based displays Gene: TCF12 (ENSG00000140262)

Orthologues Orthologues help Paralogues *

The following gene(s) have been identified as putative orthologues:
 (N.B. If you don't find a homologue here, it may be a "between-species parologue". Please view the [gene tree info](#) to see more.)

Species	Type	dn/dS	Ensembl identifier	External ref.
Alpaca (<i>Vicugna pacos</i>)	1-to-1	na	ENSPAG000000005545	TCF12 Target Nid: 98; Query Nid: 32 Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) [Source: UniProtKB/Swiss-Prot; acc: 099081]
Anole Lizard (<i>Anolis carolinensis</i>)	1-to-1	na	ENSACAG00000014277	TCF12 Target Nid: 76; Query Nid: 79 Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) [Source: UniProtKB/Swiss-Prot; acc: 099081]
Armadillo (<i>Dasyurus novemcinctus</i>)	1-to-1	0.09408	ENSDDNG00000013864	TCF12 Target Nid: 60; Query Nid: 38 Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) [Source: UniProtKB/Swiss-Prot; acc: 099081]
Bushbaby (<i>Otolemur garnettii</i>)	1-to-1	na	ENSGAG000000004595	TCF12 Target Nid: 76; Query Nid: 63 Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) [Source: UniProtKB/Swiss-Prot; acc: 099081]
Ciona savignyi	1-to-many	na	ENSCSAVVG000000011705	Novel Ensembl prediction Target Nid: 23; Query Nid: 22 No description [Multi-species view] [Aliog]
Caenorhabditis elegans	1-to-many	na	M0385.2	hlf-2 Target Nid: 23; Query Nid: 13 hlf-2 encodes a Class I basic helix-loop-helix (bHLH) transcription factor that is the <i>C. elegans</i> ortholog of the <i>Ae3</i> -acute-scuta homologs, LIN-32, a neural-specific protein with which it functions in male tail development and HLF-3, with which it is coexpressed in the nerve ring of embryonic and larval neurons and with which it regulates the expression of <i>lin-32</i> . It can be seen localized in the NGM sensory cells, in neurons, and in the nerve ring. HLF-2 is required for bestow proAC competence on the cells that undergo the AC/VU (anchor cell/ventral uterine precursor) cell fate decision, for specification, differentiation, and function of the distal tip cell (DTC), and AC, including transcriptional regulation of the LAZ-2 Delta-like ligand in the latter, and for formation of the uterine seam cell (usc). genetic analysis also suggests that HLF-2 functions with HLF-14, an additional Ae3-acute-scuta homolog, to specify the PGL-1 postembryonic stage. In addition, hlf-2 is expressed in the nerve ring of the adult hermaphrodite, where, when expression becomes increasingly restricted to neuronal cells and their immediate precursors, later expression is detected in, but not limited to, pharyngeal cells, anterior neurons, vulval and uterine muscles, the DTCs, the presumptive and mature AC, the Q neuroblast, and enteric muscles. comparative analysis of transcriptional and translational reporters indicates that hlf-2 is expressed in both the anchor cell and the ventral uterine (VU) precursor, but that expression in the latter is subject to post-transcriptional down-regulation. HLF-2 accumulation in the presynaptic AC is the first detectable difference between the AC and VU precursors during the lateral specification event that distinguishes these two cell fates. [Source: WormBase]
Cat (<i>Felis catus</i>)	1-to-1	na	ENSFCAOG000000001867	TCF12 Target Nid: 78; Query Nid: 67 Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) [Source: UniProtKB/Swiss-Prot; acc: 099081]

Ensembl genome browser S6: *H.sapiens* – Variation Image – Gene: TCF12 (ENSG00000140262)

Location: Chromosome 15: 57,210,823-57,582,051, forward strand.

Transcripts

There are 5 transcripts in the gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
TCF12-001	ENST00000267811	ENSP00000267811	protein_coding
TCF12-002	ENST00000331725	ENSP00000331057	protein_coding
TCF12-201	ENST00000343827	ENSP00000342459	protein_coding
TCF12-202	ENST00000438423	ENSP00000388940	protein_coding
TCF12-203	ENST00000452095	ENSP00000396881	protein_coding

Transcript and Gene level displays

In Ensembl a gene is made up of one or more transcripts. We provide displays at two levels:

- Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
- Gene views which provide displays for data associated at the gene level such as orthologues and paralogues, regulatory regions and splice variants.

This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.

Variation Table **Variation Image** **External Data**

Ensembl

Transcript tab

Ensembl Transcript tab: Transcript summary

Transcript: TCF12-001 (ENST00000267811)
 Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) [Source:UniProtKB/Swiss-Prot;Acc:Q99081]
 Chromosome 15: 57,210,823-57,582,051 forward strand.
 This transcript is a product of gene ENSG00000140262 - There are 5 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
TCF12-001	ENST00000267811	ENSP00000267811	protein_coding
TCF12-002	ENST00000333725	ENSP00000331057	protein_coding
TCF12-201	ENST00000343827	ENSP00000342459	protein_coding
TCF12-202	ENST00000438423	ENSP00000388940	protein_coding
TCF12-203	ENST00000452095	ENSP00000396881	protein_coding

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

Transcript summary [help](#) [Supporting evidence »](#)



Statistics Exons: 20 Transcript length: 6,061 bps Translation length: 682 residues
CCDS This transcript is a member of the Human CCDS set: [CCDS10159](#)
Type Known protein coding

Prediction Method Transcript where the Ensembl genebuild transcript and the Vega manual annotation have the same sequence, for every base pair. See [article](#).
Alternative transcripts This Ensembl/Havana merge gene entry corresponds to the following database identifiers:
 Transcript having exact match between ENSEMBL and HAVANA: [OTTHUH00000255069](#) [view all locations]

Ensembl release 56 - Sept 2009 © [\[link\]](#) [\[link\]](#) [About Ensembl](#) | [Contact Us](#) | [Help](#)
[Permanent link](#) - View in archive site

Ensembl Transcript tab: Supporting evidence

Transcript: TCF12-001 (ENST00000267811)
 Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) [Source:UniProtKB/Swiss-Prot;Acc:Q99081]
 Chromosome 15: 57,210,823-57,582,051 forward strand.
 This transcript is a product of gene ENSG00000140262 - There are 5 transcripts in this gene: [hide transcripts](#)

Name	Transcript ID	Protein ID	Description
TCF12-001	ENST00000267811	ENSP00000267811	protein_coding
TCF12-002	ENST00000333725	ENSP00000331057	protein_coding
TCF12-201	ENST00000343827	ENSP00000342459	protein_coding
TCF12-202	ENST00000438423	ENSP00000388940	protein_coding
TCF12-203	ENST00000452095	ENSP00000396881	protein_coding

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

Transcript summary [help](#) [Supporting evidence »](#) [Exons »](#)

Ensembl/Havana merge gene entry

Transcript evidence

Source	Start	End
NM_003205.3	57,210,823	57,582,051
CCDS10159.1	57,210,823	57,582,051
Exon evidence		
ALB31981.2	57,210,823	57,582,051
BC051769.2	57,210,823	57,582,051
BC051770.2	57,210,823	57,582,051
Q86TC1.1	57,210,823	57,582,051
Q86WQ2.1	57,210,823	57,582,051
Q99081.1	57,210,823	57,582,051
M86850.1	57,210,823	57,582,051
BH449501.1	57,210,823	57,582,051
BQ110370.1	57,210,823	57,582,051
BQ000505.1	57,210,823	57,582,051
BH578688.1	57,210,823	57,582,051
BH470867.1	57,210,823	57,582,051
BH702070.1	57,210,823	57,582,051
CA422291.1	57,210,823	57,582,051
BO010518.1	57,210,823	57,582,051
BH755871.1	57,210,823	57,582,051
AA049011.1	57,210,823	57,582,051
AV21359.1	57,210,823	57,582,051
AA81132.1	57,210,823	57,582,051
AU69491.1	57,210,823	57,582,051
BU178185.1	57,210,823	57,582,051
AK322720.1	57,210,823	57,582,051
NM_207037.1	57,210,823	57,582,051

Ensembl Transcript tab: Protein sequence

The screenshot shows the Ensembl transcript page for TCF12-001. The main content area displays the protein sequence:

```

M8PQQQRQAAALCCTTKEKLSCLDF3AEMFSPPNNSGPERPPTLQSSQFSGS21DEBQDT2M
CTSGPQYDPIVYQKQHCPYQTCRPNVSLNQDQGQDQGQDQGQDQGQDQGQDQGQDQGQDQGQDQGQ
SRDTGLPQCGQSLRQDGQGLGQGQPLAQGQGQPLAQGQGQPLAQGQGQPLAQGQGQPLAQGQGQ
KKVAKVNPQPGCLPSVYAPSPMSDCNFQESPFSTPFPKPFISNFASFTNQGQGTQNSQSLWSS
SMGMSQGPFGGLGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQ
STSSSEPTVVAASLTFPTPITDSDTQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQGQ
SPPVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSV
PSTSLPAGQGQDNEQGQGQDNEQGQGQDNEQGQGQDNEQGQGQDNEQGQGQDNEQGQGQDNEQGQG
SVLAVSTVITSTBLNQHKTQENYRQGQDQGQGTVVTELEKTNKEKECNLHEPPSSQSCMRS
DQESQKQDQKIVSRSGRQTSSTIMEEDCNPQKIEERKEEDRMANNADMERLRVQDNEAFAKEL
GRMCQQLLQKSEPKPTKLL1LQMWAII1SLQQQVVERNLMPAACLKRREEEKVYSAVSAS
PTTILPQTPGQPLGQTTTQPMHM

```

Annotations above the sequence include:

- Exon alternating test colour
- Residue overlap splice site
- Insert / deletion (Mouse over alternative residues)
- Synonymous SNP (Mouse over alternative residues)
- Non-synonymous SNP (Mouse over alternative residues)

The left sidebar includes a transcript-based displays menu and a configuration section for the transcript level view.

Ensembl archive

The screenshot shows the Ensembl archive page for TCF12-001. The main content area lists available archive versions:

- Ensembl 55: Jul 2009
- Ensembl 54: May 2009
- Ensembl 53: Mar 2009
- Ensembl 52: Dec 2008
- Ensembl 51: Nov 2008
- Ensembl 50: Jul 2008
- Ensembl 49: Mar 2008
- Ensembl 48: Dec 2007
- Ensembl 47: Oct 2007
- Ensembl 46: Aug 2007
- Ensembl 45: Jun 2007
- Ensembl 44: Apr 2007
- Ensembl 43: Feb 2007
- Ensembl 42: Dec 2006
- Ensembl 41: Oct 2006
- Ensembl 40: Aug 2006
- Ensembl 39: Jun 2006
- Ensembl 38: Apr 2006
- Ensembl 37: Feb 2006
- Ensembl 35: Nov 2005
- Ensembl 34: Oct 2005
- Ensembl 31: May 2005
- Ensembl 25: Oct 2004

The left sidebar includes a transcript-based displays menu and a configuration section for the archive view.

The screenshot shows the Ensembl archive protein sequence page for TCF12-001. The URL is http://dec2008.archive.ensembl.org/Homo_sapiens/Transcript/Sequence_Protein?db=core&g=ENSG000000267811. The main content includes:

- Transcript-based displays:** Transcript summary, Exons (20), Supporting evidence (24), Sequence, cDNA, Protein.
- Gene:** TCF12-001 (ENST000000267811) is a product of gene ENSG000000140262, which has 3 transcripts.
- Protein sequence:** The sequence is shown with annotations: Exon alternating text colour, Residue overlap splice site, Insert / deletion (Mouse over shows In-dele), Non-synonymous SNP (Mouse over alternative codon), and Synonymous SNP (Mouse over alternative codon).
- General identifiers:** UniProtKB/Swiss-Prot ID Q99081.

This slide is a placeholder for demonstrating how to find a chicken homolog of a human protein using Ensembl. It features the Ensembl logo and the text: "Find a chicken homolog of a human protein".

BlastView

Ensembl BLAST search

Important Notice
 We now used Blat as our default DNA search. This will make your query faster.

Enter the Query Sequence
 Either Paste sequences (max 30 sequences) in FASTA or plain text:

```
>gi|51743308|ref|NP_001455.3| ADAM metallopeptidase domain 14
MNYVPLLSCLGQLRNDNSNTDGLPQVITYPEKLRSLREKIDISQASYKIVIEKPKYI
RTYTSYSCTC1IMEPPLDQFQNMCHVYQF186PFSVWVNSTCYLGRGVQLEFENSYOII
YQVHHRKAADVSYNEEDIEEROLSLFKLQSVPEPQDNFTATILEMEVIVEKQLYNEMQSDI
```

 Or Upload a file containing one or more FASTA sequences

 Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)

 Or Enter an existing ticket ID:

 dna queries
 peptide queries

Select the databases to search against
 Select species:
 Use 'ctrl' key to select multiple species
 Citrus_galus
 Canis_aculeatus
 Corolla_gorilla
 dna database
 peptide database
 LATESTCDB
 PEP_ALL

Select the Search Tool
 BLASTN

Search sensitivity:
 Optimise search parameters to find the following alignments

About BlastView

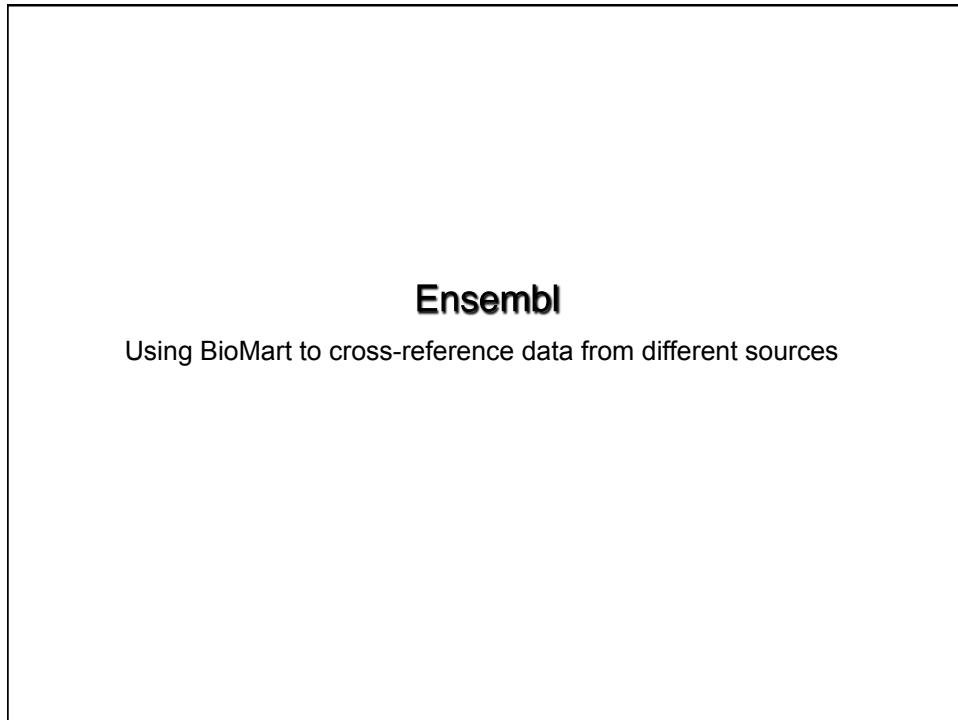
Alignment Summary (click arrow to hide)

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

refresh

Ensembl BLAST search

Query	Subject	Chromosome	Supercontig	Contig	Stats	Sort By
.off. Name Start	.off. Name Start	.off. Name Start	.off. Name Start	.off. Name Start	.off. Score E-val	>Contig <Score >Score
Links	Query	Chromosome	Supercontig	Contig	Stats	
	Start End Ori	Name				
	Start	End	Ori	Name	Start	
[A] [S] [G] [C]	4 669 +	Chr:5	29718636	29720642	+ 1465 1.e-129 33.00 708	
[A] [S] [G] [C]	6 505 +	Chr:15	6293553	6295064	+ 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]	136 515 +	Chr:22	2490093	2490449	+ 300 1.e-10 38.00 13	
[A] [S] [G] [C]	399 514 +	Chr:22	2501194	2501499	+ 300 3.1e-78 38.13 115	
[A] [S] [G] [C]	362 644 +	Chr:17	3075328	3076148	295 5.5e-21 29.28 321	
[A] [S] [G] [C]	436 570 +	Chr:22	2490093	2490449	+ 293 2.2e-70 36.36 143	
[A] [S] [G] [C]	425 659 +	Chr:1	2004	Query location : ref NP_001455.3	4 to 669 (+)	
[A] [S] [G] [C]	445 505 +	Chr:22	2471	Database location : 5	29718636 to 29720642 (+)	
[A] [S] [G] [C]	445 535 +	Chr:6	3441	Genomic location : 5	29718636 to 29720642 (+)	
[A] [S] [G] [C]	445 502 +	Chr:6	3291			
[A] [S] [G] [C]	212 270 +	Chr:22	2504	Alignment score : 1465		
[A] [S] [G] [C]	444 504 +	Chr:4	931	E-value : 1.e-129		
[A] [S] [G] [C]	339 404 +	Chr:22	2509	Alignment length : 708		
[A] [S] [G] [C]	444 501 +	Chr:22	1061	Percentage identity: 33.05		
[A] [S] [G] [C]	329 456 +	Chr:6	1041			
[A] [S] [G] [C]	349 406 +	Chr:22	2481	Query: 4 VLFPLSLGCLMDMSNPDSLPLVQIY---UPEKIRSIKREGIESQASQYKIVIEKPKYTV 58		
[A] [S] [G] [C]	407 569 +	Chr:6	1041	VL +1 GL G + *S P++T+ VP ++ S + * SX + *EG+P +		
[A] [S] [G] [C]	329 408 +	Chr:22	2471	Sbjct: 29718636 VLVLVLLGLVGCPTTPDEGSPLVHGNWTVPROL-SPRAD7NPLTVSYWLQVEGRGPQVL 29718812		
				Query: 59 S-NQFLPLPHNFNPVSYSSGTGIMPLQDF-QNFQHYQOGYIEGYPKPSVNVWSTC-TGLR 115		
				L +K F +Y G + *Q + Q+ C YQQ +G P +S+V +TC GLR		
				Sbjct: 29718813 APRPAKGKLASRPTLTYDDEGARNE-EQVYVOUNCNPYQGBEVQGSPLSGULGTGCRGLR 29718989		
				Query: 116 GVLQFENVSYGIEPLESSVGFEHVIYQVKHHKAADVSYLYNEK-DISSLSDLFSK-----LQ 168		
				GVL E +Y IED+ F+H++Y++ AD + *L ++ LQ		
				Sbjct: 29718990 GVLWNEGSTEYEIPDPPAFQFMLYRPMWQ---ADSDPMPGPCTCGLPTELEYQKTYVLPMLQ 29719160		
				Query: 169 S--VFPQ---QDF---ANV1ENMUVVKEQLYNNHNGSD---TIVVAVQVFPQLIGLNAIFV 217		
				E +D +Y+++ V+A+ + SD + V O V +--+ +++		
				Sbjct: 29719161 APRTEDKYMILKDWHTHTTRYKLVVVVNVRF--VRSDRNESKVLMQ-VLEVNVIGDSLYD 29719331		
				Query: 218 SFNITIILSLWLDENIKAATTGEANELHFTLWRKTKSYLVLR-PHDVAFLVLYRE-K 274		
				++ + L LE++ N I T A++ L F RW S L R HD A L ++ K		
				Sbjct: 29719332 QLGVQLFLVGLEWIWNNSPINITRKSASKTLADPNRNRKSDLYPRMQUDTAEFLPFGFGK 29719511		
				Query: 275 SNVVGATPGQGMCDANYAGVVLHP--RTISLESLAVILAQLLSSLNGMTYDINKQCQSG 333		
				S +G +C +D +V + R + S V + L +G+C +D+C		
				Sbjct: 29719512 8-LGLAYLGICDROWSAANDSYTNRRLS--SFIVTFVHELGHNLMRMRDE-RHKCKRR 29719676		
				Query: 334 AVC1H-NPEAIHFSGVKIFENCGCFDPDAHFESKOKSQCLRNQPRLDFFF--KQQAVCORN 390		
				CIN E S FS+C+S+D+ + + S CL+ P +R ++ K++ CON		
				Sbjct: 29719677 KRCINIESE---SDTDAFSQCCYDYPDLLGNGGS-CLYQAPALGSYTTLMHE-YCGNK 29719838		
				Query: 391 KLEAGEECCDCGTCDDCALIGETCCDIATCRPKAGSNCAEGPCCNCLFMSKERNCRCPSPF 450		
				+E+GE+CDG++ DC + CC C AGS CA G CC+ C + +CR		
				Sbjct: 29719839 IVESGECCDGSKS8DCRR--DPCTH-TAGSVCAKGCKCCKQCLPAGTLCARTG 29720009		



The screenshot displays the Ensembl BioMart interface. On the left, a sidebar lists datasets: Macaca mulatta genes (MMUL_1.0), Ensembl Gene ID, and Ensembl Transcript ID. The 'Dataset' section is currently selected, showing 'Macaca mulatta genes (MMUL_1.0)' and '[None selected]' under 'Attributes'. The main panel is titled 'Ensembl BioMart' and contains two sections:

- Step 1: Select Dataset**: Shows the dataset dropdown set to 'Ensembl 56' and the species dropdown set to 'Macaca mulatta genes (MMUL_1.0)'. A message says 'Please restrict your query using criteria below'.
- Step 2: Select Filters (input)**: Contains sections for 'REGION', 'GENE', and 'TRANSCRIPT'. Under 'GENE', there is a 'Limit to genes ...' checkbox, a 'with WikiGene ID(s)' dropdown, and radio buttons for 'Only' and 'Excluded'. A scrollable list of Ensembl Gene IDs is shown, including ENSMUG0000002226, ENSMUG0000007556, ENSMUG0000031495, ENSMUG0000020431, and ENSMUG0000006466. Other filter options include 'ID list limit', 'Transcript count >=', and 'Gene type' (miRNA, misc_RNA, Mt_rRNA, Mt_tRNA, protein_coding).

Ensembl BioMart

Please select columns to be included in the output and hit 'Results when ready'.

Dataset
 Macaca mulatta genes (MMUL_1.0)

Filters
 Ensembl Gene ID(s): [ID-list specified]

Attributes

- Ensembl Gene ID
- Ensembl Transcript ID
- Ensembl Protein ID
- Canonical transcript stable ID(s)
- Description
- Chromosome Name
- Gene Start (bp)
- Gene End (bp)
- Strand
- Associated Gene Name

Dataset
 [None Selected]

Ensembl

- Ensembl Gene ID
- Ensembl Transcript ID
- Ensembl Protein ID
- Canonical transcript stable ID(s)
- Description
- Chromosome Name
- Gene Start (bp)
- Gene End (bp)
- Strand
- Band
- Transcript Start (bp)
- Transcript End (bp)

EXTERNAL:

- Associated Gene Name
- Associated Transcript Name
- Associated Gene DB
- Associated Transcript DB
- Transcript count
- % GC content
- Gene Biotype
- Transcript Biotype
- Source
- Status (gene)
- Status (transcript)

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

External References (max 3)

- EMBL (Genbank) ID
- EntrezGene ID
- HGNC automatic gene name
- HGNC curated gene name
- miRBase Accession(s)
- miRBase ID(s)
- PDB ID
- Protein ID
- RefSeq DNA ID
- RefSeq Predicted DNA ID
- RefSeq Protein ID
- RefSeq Predicted Protein ID
- Rfam ID
- Unigene ID
- UniProt/TREMBL Accession
- UniProt/SwissProt ID
- UniProt/SwissProt Accession
- WikiGene name
- WikiGene description

Ensembl BioMart

Mozilla Firefox

Back Forward Reload Stop Home http://uswest.ensembl.org/biomart/mview/9a45c7168f3a9ac54a91927aae41e69/9a45c7168f

Dataset
 Macaca mulatta genes (MMUL_1.0)

Filters
 Ensembl Gene ID(s): [ID-list specified]

Attributes

- Ensembl Gene ID
- Ensembl Transcript ID
- Ensembl Protein ID
- Canonical transcript stable ID(s)
- Description
- Chromosome Name
- Gene Start (bp)
- Gene End (bp)
- Strand
- Associated Gene Name

Dataset
 [None Selected]

Export all results to File TSV Unique results only Go

Email notification to

View 50 rows as HTML Unique results only

Ensembl Gene ID	Ensembl Transcript ID	RefSeq Predicted DNA	Chromosome Name	Gene Start (bp)	Gene End (bp)	Strand	Associated Gene Name
ENSMU00000002226	ENSMU000000003151		11	112393309	112541929	-1	ATXN2
ENSMU00000002226	ENSMU00000003152		11	112393309	112541929	-1	ATXN2
ENSMU00000002226	ENSMU0000045594	XR_013602	11	112393309	112541929	-1	ATXN2
ENSMU00000002226	ENSMU0000045593		11	112393309	112541929	-1	ATXN2
ENSMU00000002226	ENSMU0000009671		11	21797300	27229132	-1	C12orf11
ENSMU00000002226	ENSMU0000009672		11	21797300	27229132	-1	C12orf11
ENSMU00000002226	ENSMU0000009673		11	21797300	27229132	-1	C12orf11
ENSMU00000002226	ENSMU0000009674		11	21797300	27229132	-1	C12orf11
ENSMU00000002226	ENSMU0000009675		11	123772129	123868559	1	KNT1C1
ENSMU00000004778	ENSMU0000046246	XM_001116656	11	56750261	56817957	1	LOC716540
ENSMU00000004778	ENSMU0000046246	XM_0011166528	11	56750261	56817957	1	LOC716540
ENSMU00000004778	ENSMU0000020730	XM_001116682	11	56750261	56817957	1	LOC716540
ENSMU00000004778	ENSMU0000020730	XM_001116681	11	56750261	56817957	1	LOC716540
ENSMU00000004778	ENSMU0000020730	XM_001116643	11	56750261	56817957	1	LOC716540
ENSMU00000004778	ENSMU0000020730	XM_001116636	11	56750261	56817957	1	LOC716540
ENSMU00000004778	ENSMU0000020730	XM_001116622	11	56750261	56817957	1	LOC716540
ENSMU00000005718	ENSMU0000022057	XM_001094971	11	46151137	46181235	-1	MLL2
ENSMU00000005718	ENSMU0000022057	XM_001114381	11	46151137	46181235	-1	MLL2
ENSMU00000005982	ENSMU0000022445	XM_001092151	11	42746314	42933258	1	ARI02
ENSMU00000005982	ENSMU0000022446		11	42746314	42933258	1	ARI02
ENSMU00000005982	ENSMU0000022447		11	42746314	42933258	1	ARI02
ENSMU00000006882	ENSMU0000023726	XM_001116683	11	62218117	62286547	1	LEM03
ENSMU000000020431	ENSMU0000028736	XM_001091077	11	16013797	16182831	-1	EPSPS
ENSMU000000020431	ENSMU0000047132		11	16013797	16182831	-1	EPSPS
ENSMU000000022313	ENSMU00000031392	XM_001088758	11	50588645	50709606	-1	ATF7
ENSMU000000022313	ENSMU0000005603		11	50588645	50709606	-1	ATF7
ENSMU000000022313	ENSMU0000005597		11	50588645	50709606	-1	ATF7
ENSMU000000031495	ENSMU0000045412		11	124531712	124532068	1	C12orf65

Ensembl BioMart

Please select columns to be included in the output and hit **Results when ready**.

Dataset: Macaca mulatta genes (MMUL_1.0)

Filters: Ensembl Gene ID(s): [ID-list specified]

Attributes:

- Ensembl Gene ID
- Ensembl Transcript ID
- Human Ensembl Gene ID
- Human Chromosome
- Human Chromosome Start (bp)
- Human Chromosome End (bp)

Step 3: Select Attributes (output)

HUMAN ORTHOLOGS:

Orthologs:

- Human Ensembl Gene ID
- Representative Protein ID
- Human Ensembl Protein ID
- Human Chromosome
- Human Chromosome Start (bp)
- Human Chromosome End (bp)
- dN
- dS
- Bootstrap/Duplication Confidence Score Type
- Bootstrap/Duplication Confidence Score
- % Identity
- Human % Identity

Ensembl Gene ID	Ensembl Transcript ID	Human Ensembl Gene ID	Human Chromosome	Human Chromosome Start (bp)	Human Chromosome End (bp)
ENSMUG00000002226	ENSMUT00000003151	ENSG00000204842	12	111890018	112037480
ENSMUG00000002226	ENSMUT00000003152	ENSG00000204842	12	111890018	112037480
ENSMUG00000002226	ENSMUT00000045594	ENSG00000204842	12	111890018	112037480
ENSMUG00000002226	ENSMUT00000045593	ENSG00000204842	12	111890018	112037480
ENSMUG00000006466	ENSMUT00000009071	ENSG00000064102	12	27058118	27091254
ENSMUG00000006466	ENSMUT00000009070	ENSG00000064102	12	27058118	27091254
ENSMUG00000006466	ENSMUT00000009072	ENSG00000064102	12	27058118	27091254
ENSMUG00000007556	ENSMUT00000010566	ENSG00000184445	12	123011809	123110947
ENSMUG00000014778	ENSMUT00000048246	ENSG00000118596	12	60083126	60175407
ENSMUG00000014778	ENSMUT00000020730	ENSG00000118596	12	60083126	60175407
ENSMUG00000015718	ENSMUT00000022067	ENSG00000167548	12	49412762	49449107
ENSMUG00000015962	ENSMUT00000023445	ENSG00000189079	12	46123492	46301823
ENSMUG00000015962	ENSMUT00000023446	ENSG00000189079	12	46123492	46301823
ENSMUG00000015962	ENSMUT00000023447	ENSG00000189079	12	46123492	46301823
ENSMUG00000016892	ENSMUT00000023728	ENSG00000174106	12	65563371	65642107

NCBI

View a genomic region between two SNPs

Map Viewer results page

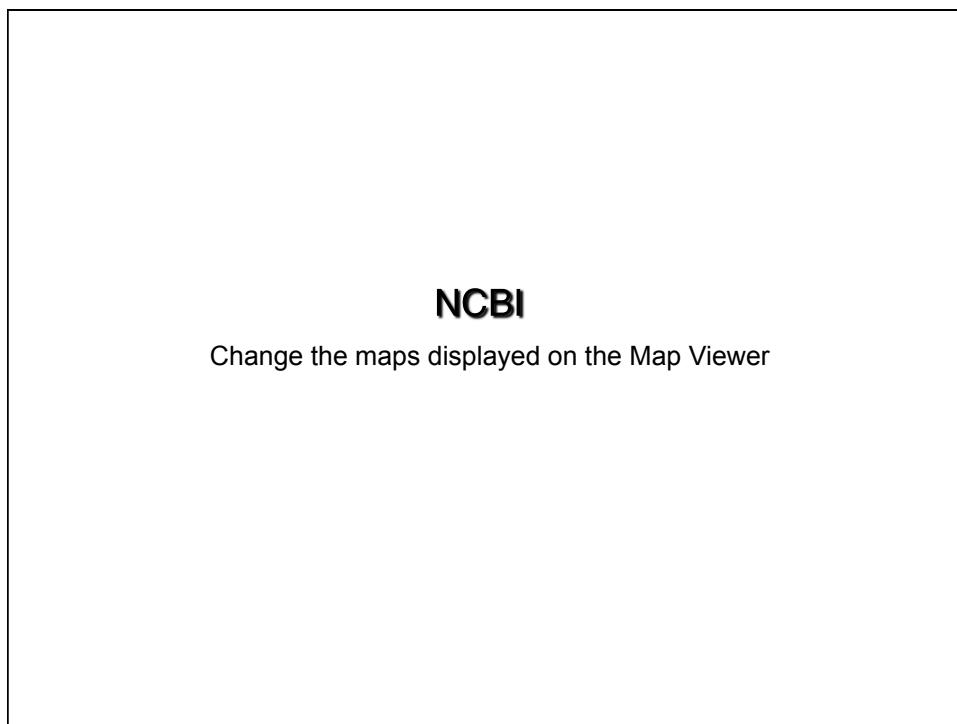
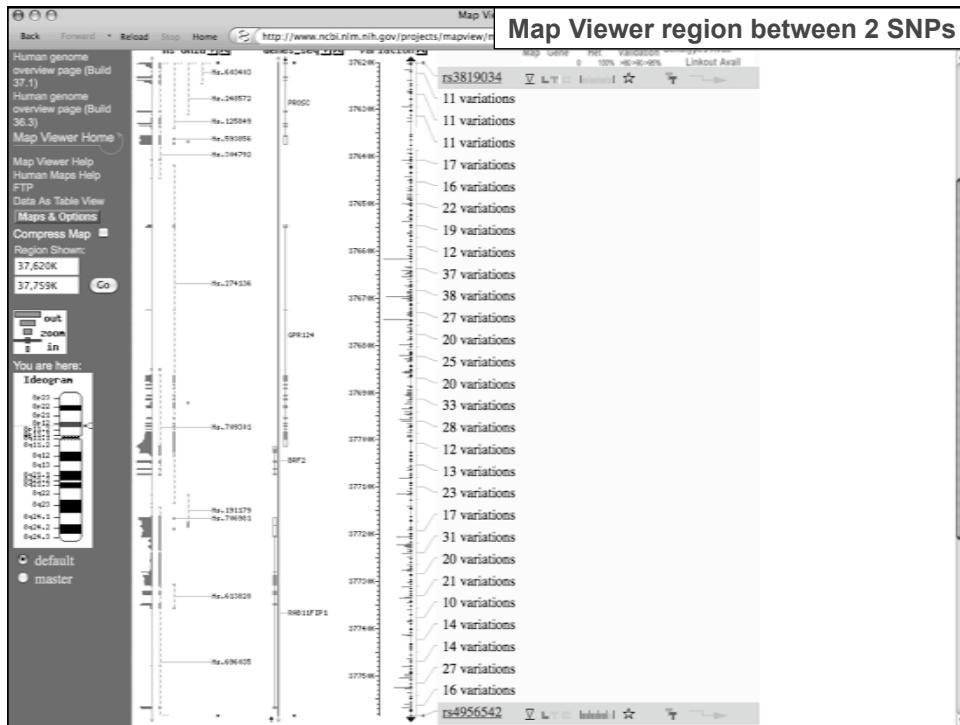
The screenshot shows a search results page for the NCBI Map Viewer. The search query is "rs3819034 OR rs4956542". The results table lists six hits across three reference builds:

Chr	Assembly	Match	Map element	Type	Maps
8	reference	all matches	rs4956542	SNP	Variation
8	Celera	all matches	rs4956542	SNP	Variation
8	HuRef	all matches	rs4956542	SNP	Variation
			rs3819034	SNP	Variation
			rs3819034	SNP	Variation

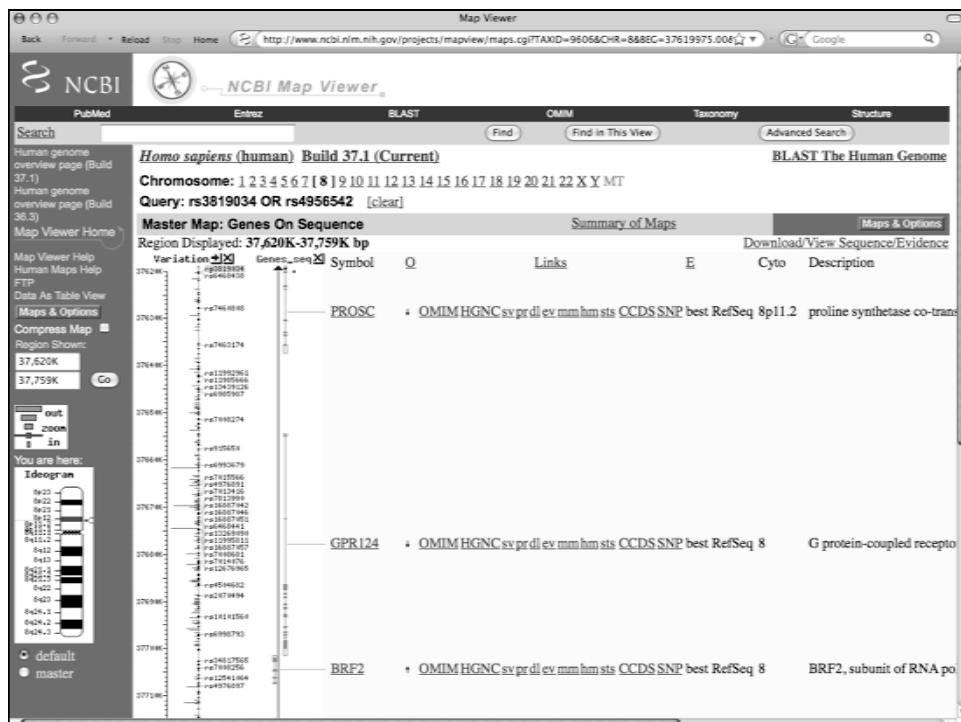
Map Viewer default view

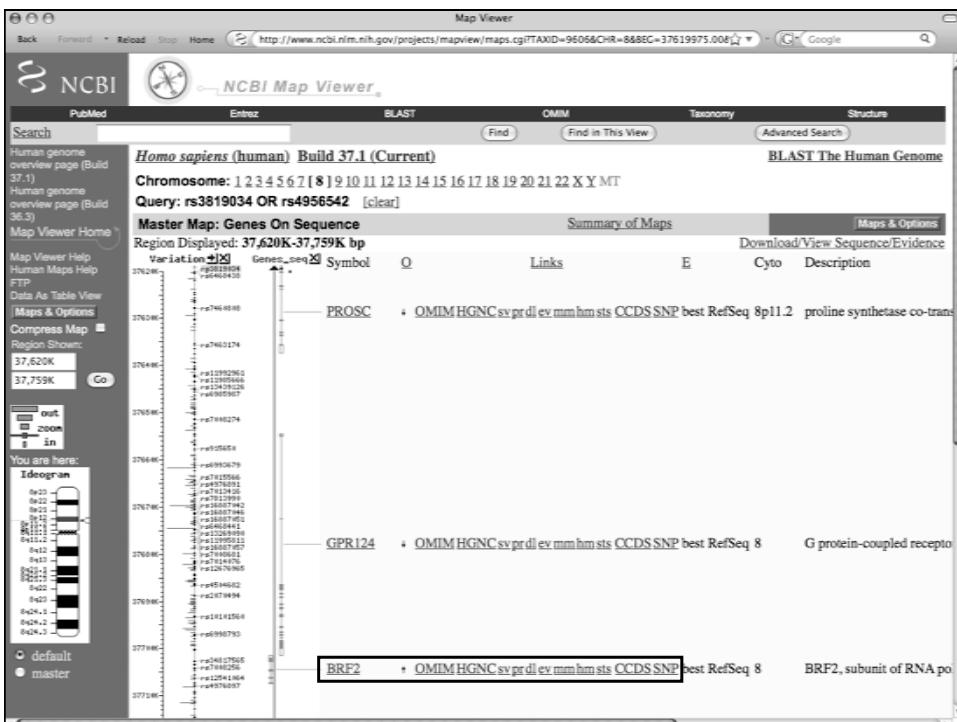
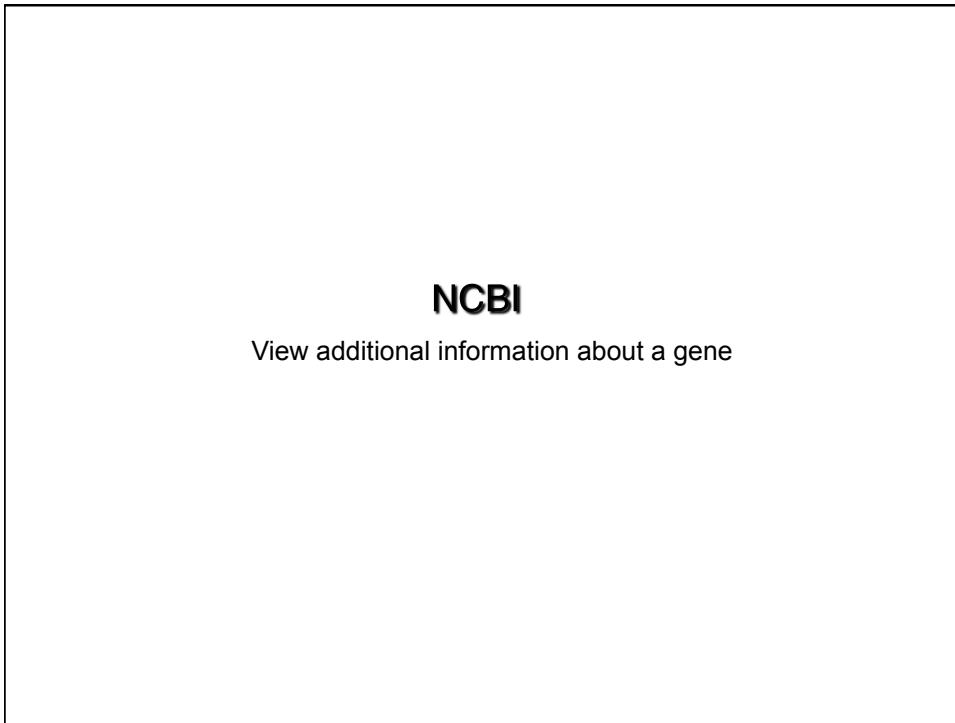
The screenshot shows the NCBI Map Viewer default view for the region 37,603K-37,776K bp. The main panel displays a detailed variation map with tracks for Hs Unit, Gene, seq, Variation, and Variation. The Variation track shows 6 hits for rs3819034 and 18 hits for rs4956542. The right side of the screen shows a summary of maps and a list of variations with their counts:

- rs3819034: 14 variations
- rs3819034: 13 variations
- rs3819034: 16 variations
- rs3819034: 18 variations
- rs3819034: 23 variations
- rs3819034: 26 variations
- rs3819034: 41 variations
- rs3819034: 41 variations
- rs3819034: 34 variations
- rs3819034: 32 variations
- rs3819034: 23 variations
- rs3819034: 42 variations
- rs3819034: 29 variations
- rs3819034: 11 variations
- rs3819034: 27 variations
- rs3819034: 8 variations
- rs3819034: 36 variations
- rs3819034: 26 variations
- rs3819034: 26 variations
- rs3819034: 12 variations
- rs3819034: 17 variations
- rs3819034: 33 variations
- rs3819034: 19 variations
- rs4956542: 18 variations



The screenshot shows the NCBI Map Viewer interface. On the left, a sidebar menu includes 'Map Viewer Home', 'Maps & Options' (which is highlighted with a black arrow and the word 'click'), and other options like 'Data As Table View'. The main window title is 'Map Viewer Maps & Options'. It displays 'Organism: Homo sapiens' and 'Chromosome: 8'. The 'Region Shown' is set to 37619975.00 - 37758538.00. The 'Available Maps' section lists various map types: Sequence Maps, Ab initio, Assembly, Celera Genes, Celera Transcripts, Clone, Component, Contig, and CpG Island. The 'Maps Displayed (left to right)' section shows a list of items: Variation, Gene, Move UP, Move DOWN, Make Master/Move to Bottom, and Toggle Ruler. Below these are 'More Options' for Show Connections, Verbose Mode, Compress Map, Auto Compress if > 350 px, Page Length (set to 30), and Thumbnail View (set to default). Buttons for OK, Apply, and Close are at the bottom.





NCBI: Entrez Gene

BRF2 BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like [Homo sapiens]

GeneID: 55290 updated 23-Jan-2010

Summary

Official Symbol BRF2 provided by HGNC

Official Full Name BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like provided by HGNC

Primary source HGNC:17298

See related Ensembl:ENSG00000104221; HPRD:06115; MIM:607013

Gene type protein coding

RefSeq status REVIEWED

Organism Homo sapiens

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchoontonglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as BRFU; FLJ11052; TFIIIB; BRF2

Summary This gene encodes one of the multiple subunits of the RNA polymerase III transcription factor complex required for transcription of genes with promoter elements upstream of the initiation site. The product of this gene, a TFIIIB-like factor, is directly recruited to the TATA-box of polymerase III small nuclear RNA gene promoters through its interaction with the TATA-binding protein. [provided by RefSeq]

Genomic regions, transcripts, and products

(minus strand) Go to reference sequence details Try our new Sequence Viewer

NC_000008.10

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)

- NM_018310.2 → NP_060780.2 RNA polymerase III transcription initiation factor BRF2**
 - Source sequence(s): AF205673, AK001914
 - Consensus CDS: CCDS6098.1
 - UniProtKB/Swiss-Prot: Q5HAW0
 - Related Ensembl: ENSP00000220659, ENST00000220659
 - Conserved Domains (2) summary

COG1405	SUA7; Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIIB [Transcription]
Locations: S - 233	Blast Score: 252
pfam08271	TFIIB, Zn_Ribbon; TFIIIB zinc-binding
Locations: S - 37	Blast Score: 112

Related Sequences

Nucleotide	Protein
genomic AC130304.12 (90409..96223)	None
genomic AC138356.3	None
genomic CH471080.2	EAW63351.1
	EAW63352.1
	EAW63353.1
mRNA AF130058.1	AAG35486.1
mRNA AF205673.2	AAG35669.2
mRNA AF298153.1	AAG30222.1
mRNA AK001914.1	BAA91975.1
mRNA AK294337.1	BAG57607.1
mRNA AK315420.1	BAG37809.1
mRNA BC010648.1	AAH10648.1
mRNA CR591369.1	None
mRNA CR602291.1	None
mRNA CR604537.1	None
mRNA CR607467.1	None
mRNA CR613791.1	None

OMIM - BRF2 SUBUNIT OF RNA POLYMERASE III TRANSCRIPTION INITIATION FACTOR; BRF2

<http://www.ncbi.nlm.nih.gov/entrez/dispmim.cgi?id=607013>

NCBI: OMIM

MIM *607013
 Cloning
 Gene Function
 Mapping
 References
 Contributors
 Creation Date
 Edit History

* Gene map
 Entrez Gene
 Nomenclature
 RefSeq
 GenBank
 Protein
 UniGene

LinkOut
 Komp
 MGI

Search OMIM for: Go Clear

Limits Preview/Index History Clipboard Details

Display Detailed Show 20 Send to

***607013**
BRF2 SUBUNIT OF RNA POLYMERASE III TRANSCRIPTION INITIATION FACTOR; BRF2

Alternative titles; symbols
BRFU
TRANSCRIPTION FACTOR IIB-RELATED FACTOR; TFIIIB50

Gene map locus 8p11.23

TEXT

CLONING

By database searching with the sequence of BRF1 (604902) as query, Schramm et al. (2000) identified BRF2, which they called BRFU, and cloned the corresponding cDNA from a total HeLa cell cDNA library. BRFU encodes a 419-amino acid protein containing a zinc ribbon domain and a core domain sharing significant homology with TFIIIB (189963) and BRF1. ☺

GENE FUNCTION

Schramm et al. (2000) found that antibodies directed against BRFU were inhibitory in an in vitro translation assay dependent upon the U6 promoter. Cabart and Murphy (2001) provided evidence that BRFU is involved in the nucleation of a polymerase III-specific small nuclear RNA (snRNA) transcription initiation complex. Using recombinant proteins and HeLa cell nuclear extracts in GST pull-down and electrophoretic gel mobility assays, they determined that BRFU does not directly bind DNA; rather, it interacts with the TATA-binding protein (TBP; 600075), and these then bind cooperatively to the polymerase III promoter region of snRNA. The TBP/BRFU complex does not appear to have strict requirements for sequence motifs outside the TATA box. With use of deletion and point mutations, they determined that arg235 and phe250 of TBP, and repeat 2 of the BRFU core domain, are necessary for complex formation with DNA. BRFU also forms a stable complex on the TATA box of the polymerase II-specific adenovirus major late promoter. ☺

HomoloGene Result

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=homologene&cmd=Link&LinkName=geneshm>

NCBI: HomoloGene (hm)

Search HomoloGene for: Go Clear

Limits Preview/Index History Clipboard Details

Display HomoloGene Show 20 Send to

All: 1 Fungi: 0 Mammals: 0

1: HomoloGene:10127. Gene conserved in Euteleostomi

Genes
 Genes identified as putative homologs of one another during the construction of HomoloGene.

Proteins
 Proteins used in sequence comparisons and their conserved domain architectures.

Download , Links

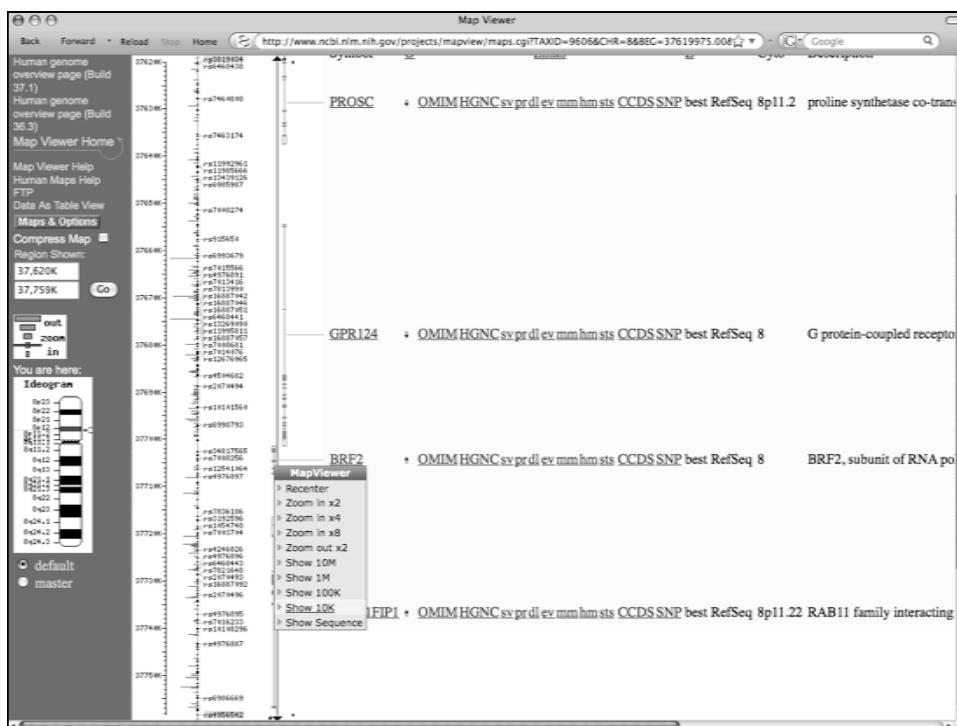
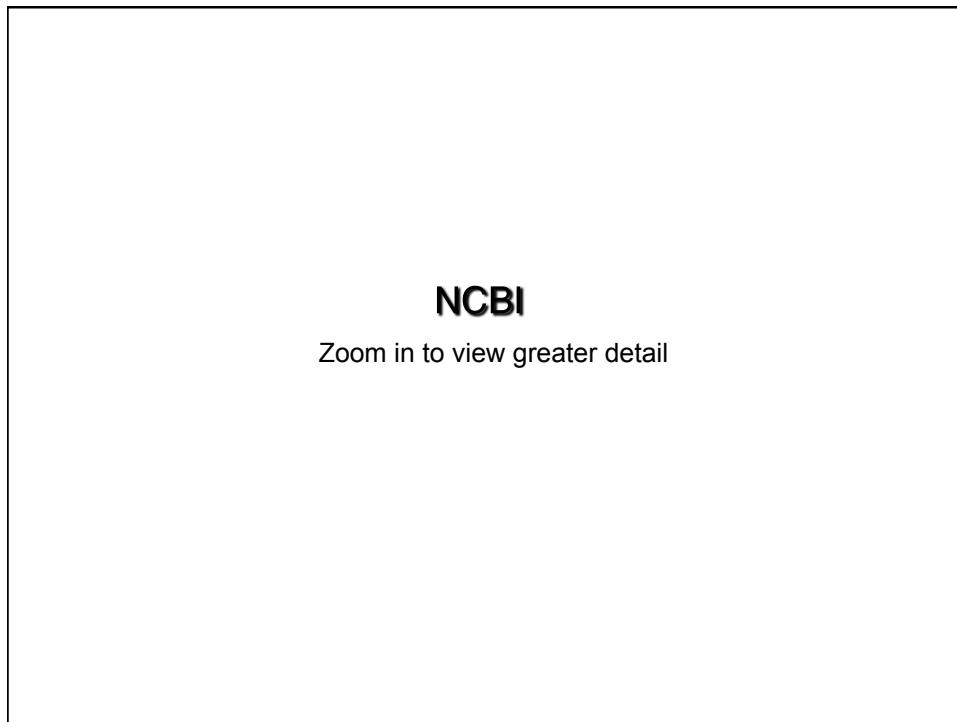
	BRF2, <i>Homo sapiens</i>	NP_060780.2 419 aa
	BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like	XP_001169914.1 419 aa
	BRF2, <i>Pan troglodytes</i>	XP_850010.1 421 aa
	BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like	NP_001015582.1 421 aa
	BRF2, <i>Canis lupus familiaris</i>	NP_079962.1 420 aa
	BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like	NP_001019644.1 416 aa
	BRF2, <i>Bos taurus</i>	XP_424383.1 415 aa
	BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like	NP_001003536.1 423 aa

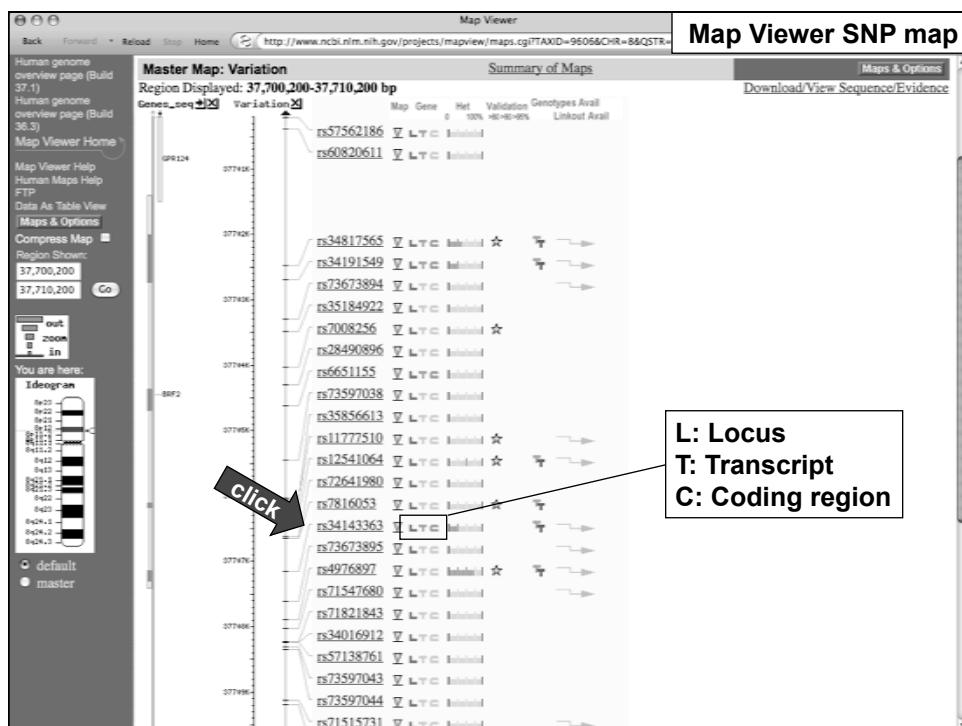
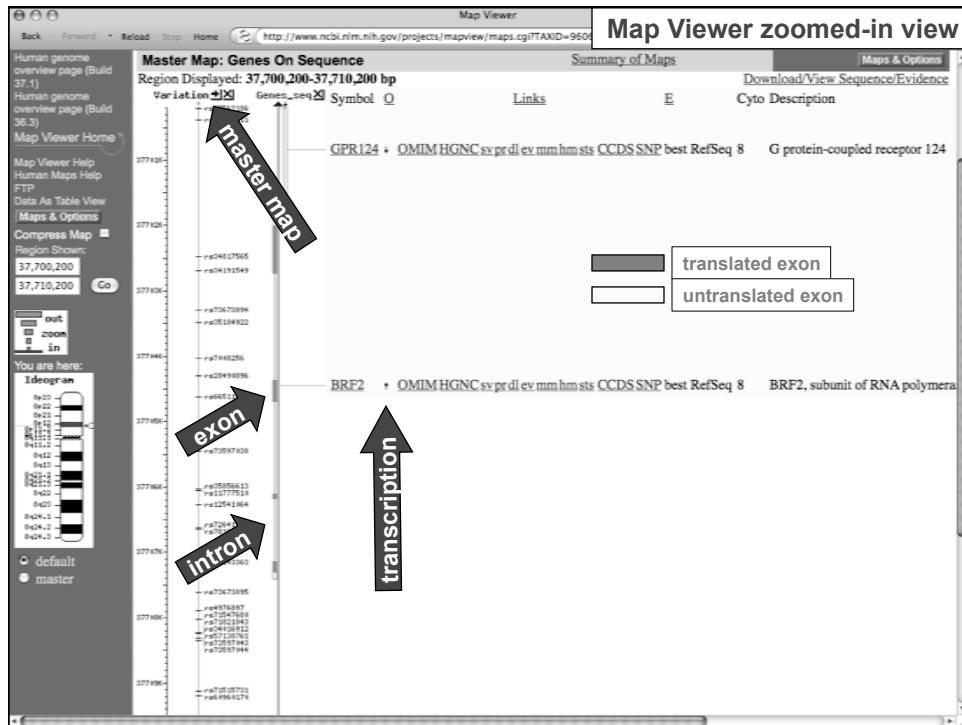
Protein Alignments
 Protein multiple alignment, pairwise similarity scores and evolutionary distances.

Conserved Domains
 Conserved Domains from CDD found in protein sequences by rpsblast searching.

Show Multiple Alignment
 Show Pairwise Alignment Scores

TFIIB_Zn_Ribbon (pfam08271)
 ■ TFIIB zinc-binding





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/MapViewHelp.html>

Current Protocols in Bioinformatics

The UCSC Genome Browser

UNIT 1.4

The rapid progress of public sequencing and mapping efforts on vertebrate genomes has increased the demand for tools that offer quick and efficient access to genomic data at all levels and facilitate comparative data analysis. The University of California Santa Cruz (UCSC) Genome Bioinformatics Web site at <http://genome.ucsc.edu> provides a wide variety of genome analysis tools, most notably the UCSC Genome Browser—supplies convenient access to the MySeq database and a collection of aligned annotation “tracks.” Another tool, the Table Browser—supplies convenient access to the MySQL database underlying the Genome Browser annotations. The custom annotation tracks feature that enables users to upload their own data and compare it against the genome.

The main protocol of this unit (see Basic Protocol) describes how to use the UCSC Genome Browser through a specific section of a genome and its annotations.

Using the NCBI Map Viewer to Browse Genomic Sequence Data

UNIT 1.5

This unit includes an introduction to the Map Viewer (see Basic Protocol), which describes how to perform a simple text-based search of genome annotations to view the results, navigate along a chromosome, zoom in and out, and change the scale and show information. It also describes some of NCBI’s other genome browsers, which are provided as links from the Map Viewer. The Alternate Protocol 1 shows how to query the genome sequence, and also illustrates how to use the Map Viewer. Alternate Protocol 1 shows how to perform a BLAST search against the human genome. Alternate Protocol 2 shows how to retrieve a list of all genes between two STS markers. Protocol 3 shows how to find all annotated members of a genome.

Using the Ensembl Genome Server to Browse Genomic Sequence Data

UNIT 1.15

The Ensembl project presents the latest sequence assembly of the human genome and provides automatic annotation of that sequence, including gene, transcript, and protein predictions. The annotation is integrated with external data sources, making Ensembl a valuable starting and reference point for any work in human biology or medicine that utilizes genetic information.

A central element of the Ensembl project is openness: all data are freely available and all the computer code used to analyze and present the data is freely available as well. More information on the Ensembl gene prediction and annotation system, and on additional ways of accessing the data, is provided in the Commentary.

This unit explains how to access and use the human sequence (although these instructions would be applicable to any of the species available in the browser) and its annotation via the Ensembl Web site. The Web site is an advanced interactive service, providing a range of views that present different aspects of the data. The Ensembl human home page (http://www.ensembl.org/Homo_sapiens) provides access to the data in several different ways, including text searches, clickable chromosomes, and sequence similarity searching in BLASTView, as well as by using the BioMart data warehouse or by simply entering chromosome coordinates.

Access through
<http://nihlibrary.nih.gov/ResearchTools/OnlineJournals.htm>