

NATIONAL HUMAN GENOME RESEARCH INSTITUTE *Division of Intramural Research*



*Current Topics in Genome Analysis
Fall 2006*

Week 4: Mining Genomic Sequence Data

Tyra G. Wolfsberg, Ph.D.

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



Accessing public genome sequence data

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

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

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

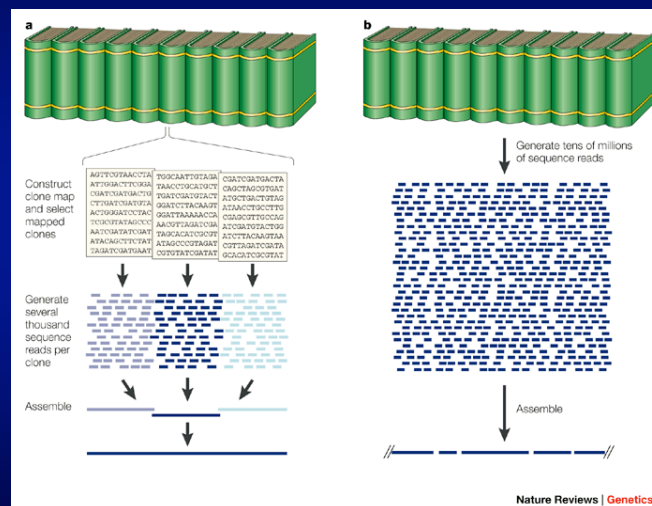
Types of data integrated in genome browsers

- 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/>
 - preEnsembl: <http://pre.ensembl.org/>
 - UCSC and Ensembl provide access to older genome assemblies and annotations; NCBI provides access only to old mouse and human data
- 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	Mar 2006/hg18/Build 36.1	Build 36.1	Build 36
Mouse	YES	Feb 2006/mm8/Build 36	Build 36.1	Build 36
Rat	YES	Nov 2004/rn4/RGSC 3.4	RGSC 3.4	RGSC 3.4
Zebrafish	NO	Mar 2006/danRer4/Zv6	Build 1.1/Zv4	Zv6
Rhesus	YES	Jan 2006/rheMac2/v.1.0, Mmul_051212	Build 1.1/v.1.0, Mmul_051212	Mmul_1
Fugu	NO	Aug 2002/ fr1/v3.0	-	Fugu 4.0

NCBI Reference Sequences (RefSeqs)

- Derived from primary GenBank submissions
- Varying levels of validation, additional annotation, and manual curation

NC_123456	Genomic	Mixed	Complete genomic molecules including genomes, chromosomes, organelles, plasmids.	NT_123456	Genomic	Automated	Intermediate genomic assemblies of BAC and/or Whole Genome Shotgun sequence data
NG_123456	Genomic	Mixed	Incomplete genomic region; supplied to support the NCBI Genome Annotation pipeline. Represents either non-transcribed pseudogenes, or larger regions representing a gene cluster that is difficult to annotate via automatic methods.	NW_123456	Genomic	Automated	Intermediate genomic assemblies of BAC or Whole Genome Shotgun sequence data
NM_123456	mRNA	Mixed	Transcript products; Mature RNA (mRNA) protein-coding transcripts.	NZ_ABCD12345678	Genomic	Automated	A collection of whole genome shotgun sequence data for a project. Accessions are not tracked between releases. The first four characters following the underscore (e.g. 'ABCD') identifies a genome project.
NM_123456789	mRNA	Mixed	Transcript products; 9-digit expansion of accession series	XM_123456	mRNA	Automated	Transcript products; model mRNA provided by the Genome Annotation process; sequence corresponds to the genomic contig.
NP_123456	Protein	Mixed	Protein products; primarily full-length precursor products but may include some partial proteins and mature peptide products.	XP_123456	Protein	Automated	Protein products; model proteins provided by the Genome Annotation process; sequence corresponds to the genomic contig.
NP_123456789	Protein	Curation	Protein products; 9-digit expansion of accession series	XR_123456	RNA	Automated	Transcript products; model non-coding transcripts provided by the Genome Annotation process.
NR_123456	RNA	Mixed	Non-coding transcripts including structural RNAs, transcribed pseudogenes, and others				

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

```

LOCUS       NM_001101                1793 bp     mRNA       linear   PRI 17-SEP-2006
DEFINITION Homo sapiens actin, beta (ACTB), mRNA.
ACCESSION  NM_001101
VERSION   NM_001101.2   GI:5016088
KEYWORDS  .
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Cranista; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
           Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1793)
AUTHORS   Pappenberger,C., McCormack,E.A. and Willison,K.R.
TITLE     Quantitative actin folding reactions using yeast CCT purified via
           an internal tag in the CCT3/gamma subunit
JOURNAL   J. Mol. Biol. 360 (2), 484-496 (2006)
PUBMED   16762366

REFERENCE  154 (bases 1 to 1793)
AUTHORS   Vandekerckhove,J. and Weber,K.
TITLE     Mammalian cytoplasmic actins are the products of at least two genes
           and differ in primary structure in at least 35 identified positions
           from skeletal muscle actins
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 75 (3), 1106-1110 (1978)
PUBMED   774701
COMMENT   REVIEWED REFSEQ: This record has been curated by NCBI staff. The
           reference sequence was derived from U00311.1 and M11312.1.
           On Jun 8, 1999 this sequence version replaced gi:4501884.

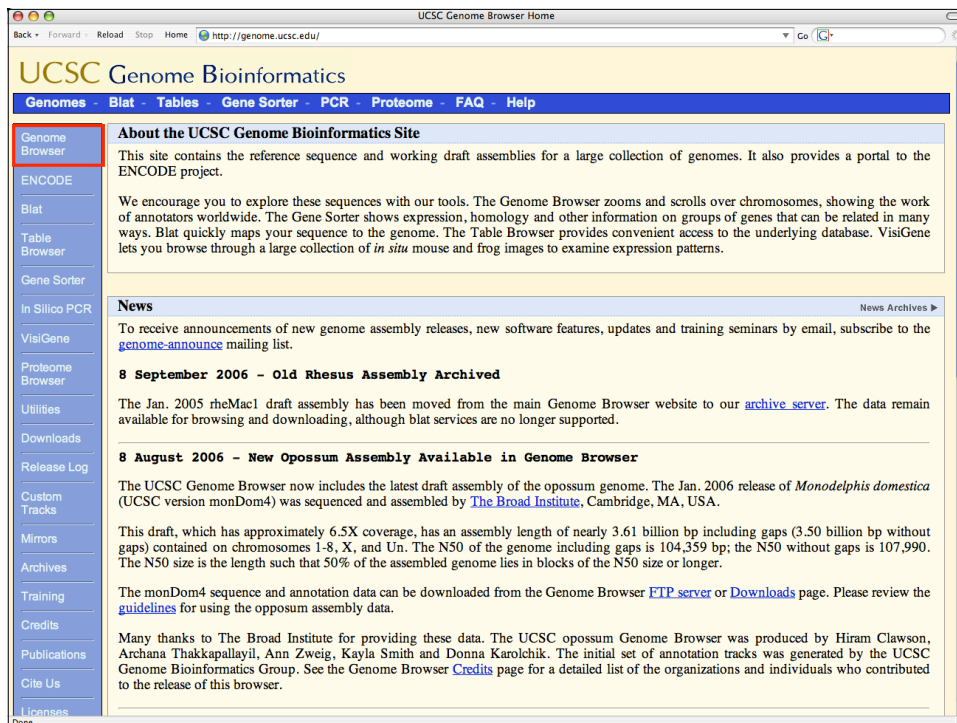
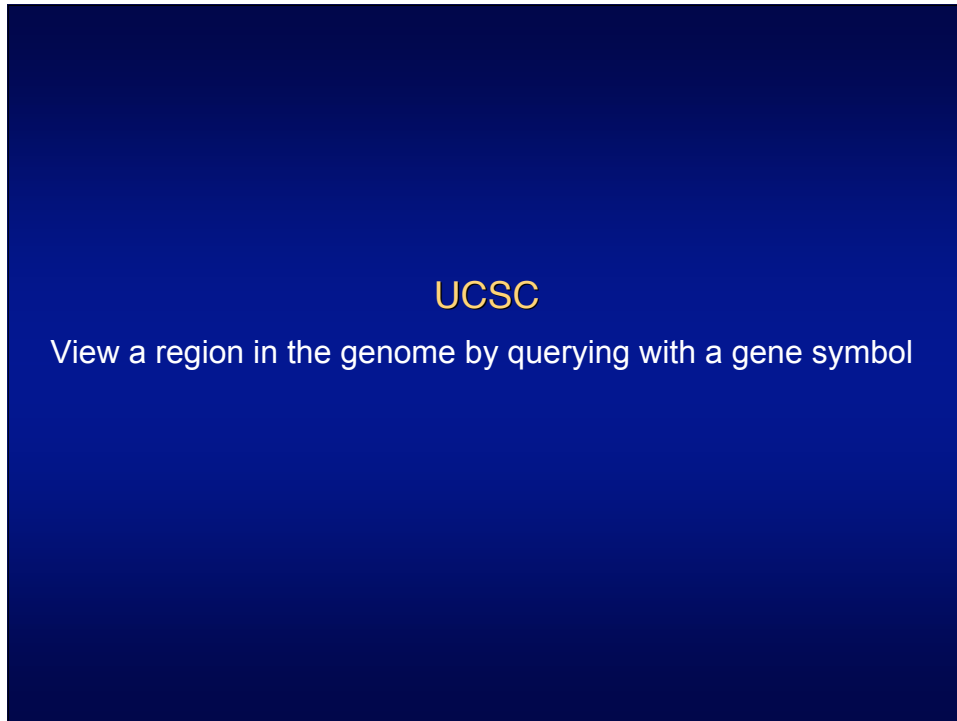
           Summary: Beta actin is one of six different actin isoforms which
           have been identified. ACTB is one of the two nonmuscle
           cytoskeletal actins. Actins are highly conserved proteins that are
           involved in cell motility, structure and integrity. Alpha actins
           are a major constituent of the contractile apparatus.
           COMPLETENESS: complete on the 3' end.

CDS       74..1201
           /gene="ACTB"
           /gc_component="actin filament; cytoskeleton; TIP60 histone
           acetyltransferase complex [pmid 1096108]"
           /gc_function="ATP binding; nucleotide binding; protein
           binding [pmid 1552781]; structural constituent of
           cytoskeleton"
           /note="beta cytoskeletal actin; PS1TP5-binding protein 1"
           /codon_start=1
           /product="beta actin"
           /protein_id="NP_001092.1"
           /db_xref="GI:4501885"
           /db_xref="CCDS:CCDS341.1"
           /db_xref="GeneID:450"
           /db_xref="HNCI:112"
           /db_xref="HPRD:0032"
           /db_xref="MIM:102610"
           /translation="MDDIALLVVDNGSMCKAGFACDDAPRAVPSIVGRPHQDVM
           VMGQKDYVGDRAQSKRGILZLVYPIEIGIVTNWDMDEKIMHFTFYNELRVAPEDIP
           VLETFAPAPANKERKFTIMETFTWPAWVAQAVLEIAPSGTGIWMSGDGVT
           HTPVIEGYALPHAILRLDLACRDLTDYLMKILTERGSPFTTAREIVROIKELCY
           VALDFQMATANSSSELEKVELPDGQVTIIGNDRFPCPALPQSFCLMESCOTHE
           FTENSINQVDIIEKLYAVYLSGCTWYICLADWQKE;TALAFKRWKIIAPPE
           RKYSVWIGGSLASLSTFQMWISKQYVDSGSPVHRKCF"

ORIGIN
1  ccgctcggcc  ccgagagcac  agagctcggc  cttggccgat  ccggcgccgc  tcaaacccy
61  ccgcccggtc  ggtctctcgg  atgatctcgc  cggcgtctgc  gtcccaagcg  gctccggcat
121  gtcgcaagcc  ggtctcggcg  ggcagcagtc  cccccggccc  gttctccctc  ceatcgtggg
181  ccgcccaagc  caccagagcg  tgatggtggg  catggctcag  agagatcctc  atgtgggca
241  ccaggcccaq  agcaagagag  gcatctctac  cctgagatcc  cccatcagag  acagatcgt
241

```

Beta actin mRNA RefSeq



Human (Homo sapiens) Genome Browser Gateway

Home Genomes Blat Tables Gene Sorter PCR FAQ Help

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: Vertebrate genome: Human assembly: Mar. 2006 position or search term: ADAM2 image width: 620 submit

Human Chimp Rhesus Dog Cow Mouse Rat Opossum Chicken X. tropicalis Zebrafish Tetraodon Fugu

Human (Homo sapiens) Genome Browser Gateway

The March 2006 human reference sequence (hg18) was produced by the International Human Genome Sequencing Consortium.

Sample position queries

A gene position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS 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
chr3:100000+2000 Displays a region of chr3 that spans 2000 bases, starting with position 1000000

D1653046 Displays region around STS marker D1653046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
RH18061;RH80175 Displays region between STS markers RH18061;RH80175. Includes 100,000 bases on each side as well. This syntax may also be used for other range queries, such as between cytobands and uniquely-determined ESTs, mRNAs, refSeqs, etc.

AA205474 Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101 Displays region of clone with GenBank accession AC008101
AF083811 Displays region of mRNA with GenBank accession number AF083811
PRNP Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414 Displays the region of genome with RefSeq identifier NM_017414
NP_059110 Displays the region of genome with protein accession number NP_059110

pseudogene mRNA Lists transcribed pseudogenes, but not cDNAs
homeobox caudal Lists mRNAs for caudal homeobox genes
zinc:finer Lists many zinc:finer mRNAs

Human ADAM2 - UCSC Genome Browser v141

Back Forward Reload Stop Home http://genome.ucsc.edu/cgi-bin/hgTracks?clade=vertebrate&org=Human&db=hg18&position=ADAM2&ix=620&hgssid=7752169

Known Genes

ADAM2 (NM_001464) at chr8:39720414-39814936 - ADAM metalloproteinase domain 2 proprotein
ADAM2 (BC064547) at chr8:39720414-39814886 - ADAM2 protein.
ADAM2 (BC034957) at chr8:39720414-39814885 - ADAM2 protein.
ADAM22 (NM_021723) at chr7:87401638-87664383 - ADAM metalloproteinase domain 22 isoform 1
ADAM22 (NM_021721) at chr7:87401638-87664364 - ADAM metalloproteinase domain 22 isoform 5
ADAM22 (NM_016351) at chr7:87401638-87664383 - ADAM metalloproteinase domain 22 isoform 3
ADAM29 (NM_014269) at chr8:176076134-176135905 - ADAM metalloproteinase domain 29 preproprotein
ADAM28 (NM_014265) at chr8:24207525-24268670 - ADAM metalloproteinase domain 28 isoform 1
ADAM22 (NM_004194) at chr7:87401638-87664364 - ADAM metalloproteinase domain 22 isoform 4
ADAM20 (NM_003814) at chr14:70058832-70061255 - ADAM metalloproteinase domain 20 preproprotein
ADAM21 (NM_003813) at chr14:69939370-69936374 - ADAM metalloproteinase domain 21 preproprotein
ADAM23 (NM_003812) at chr2:207016613-207190222 - ADAM metalloproteinase domain 23 preproprotein
ADAM20 (AF029899) at chr14:70058832-70061255 - ADAM metalloproteinase domain 20
ADAM22 (BC036022) at chr7:87401638-87664383 - ADAM22 protein.
ADAM22 (AF155381) at chr7:87401638-87663791 - Hypothetical protein ADAM22 (Fragment).
ADAM18 (AY358321) at chr8:139561257-139706740 - ADAM metalloproteinase domain 18

RefSeq Genes

ADAM2 at chr8:39720414-39814936 - (NM_001464) ADAM metalloproteinase domain 2 proprotein
ADAM20 at chr14:70058832-70061255 - (NM_003814) ADAM metalloproteinase domain 20 preproprotein
ADAM21 at chr14:69939370-69936374 - (NM_003813) ADAM metalloproteinase domain 21 preproprotein
ADAM22 at chr7:87401638-87664364 - (NM_004194) ADAM metalloproteinase domain 22 isoform 4
ADAM22 at chr7:87401638-87664364 - (NM_021721) ADAM metalloproteinase domain 22 isoform 5
ADAM22 at chr7:87401638-87664383 - (NM_021722) ADAM metalloproteinase domain 22 isoform 2
ADAM22 at chr7:87401638-87664383 - (NM_021723) ADAM metalloproteinase domain 22 isoform 1
ADAM22 at chr7:87401638-87664383 - (NM_016351) ADAM metalloproteinase domain 22 isoform 3
ADAM28 at chr8:24207525-24268670 - (NM_021777) ADAM metalloproteinase domain 28 isoform 3
ADAM28 at chr8:24207525-24268670 - (NM_014265) ADAM metalloproteinase domain 28 isoform 1
ADAM29 at chr8:176076134-176135905 - (NM_014269) ADAM metalloproteinase domain 29 preproprotein

Non-Human RefSeq Genes

ADAM2 at chr8:39720414-39814936 - (NM_213957) fertilin beta
Adam2 at chr8:39720414-39813888 - (NM_17428) a disintegrin and metalloproteinase domain 2
Adam2 at chr8:39720414-3981569 - (NM_009618) a disintegrin and metalloproteinase domain 2
Adam2 at chr8:39720414-39813877 - (NM_020077) a disintegrin and metalloproteinase domain 2
Adam21 at chr14:69939370-69936374 - (NM_020330) a disintegrin and metalloproteinase domain 21
Adam21 at chr14:69781958-70061197 - (NM_020330) a disintegrin and metalloproteinase domain 21
Adam22 at chr7:87401638-87664383 - (NM_001007220) a disintegrin and metalloproteinase domain 22
Adam22 at chr7:87401638-87664383 - (NM_001007221) a disintegrin and metalloproteinase domain 22
Adam23 at chr2:207016613-207190222 - (NM_011780) a disintegrin and metalloproteinase domain 23
Adam23 predicted at chr2:207016613-207190222 - (NM_001029899) a disintegrin and metalloproteinase domain 23
Adam24 at chr9:17371373-17373360 - (NM_010086) a disintegrin and metalloproteinase domain 24
Adam24 at chr9:18904670-18906353 - (NM_010086) a disintegrin and metalloproteinase domain 24
Adam24 at chr14:70059197-70060969 - (NM_010086) a disintegrin and metalloproteinase domain 24
Adam25 at chr1:18904411-18906248 - (NM_011781) a disintegrin and metalloproteinase domain 25
Adam25 at chr1:18904670-18906353 - (NM_011781) a disintegrin and metalloproteinase domain 25
Adam25 at chr14:6994072-6995994 - (NM_011781) a disintegrin and metalloproteinase domain 25
Adam25 at chr14:6978108-69783962 - (NM_011781) a disintegrin and metalloproteinase domain 25
Adam26 at chr3:17171523-17173287 - (NM_010085) a disintegrin and metalloproteinase domain 26
Adam26 at chr3:17171523-17173287 - (NM_010085) a disintegrin and metalloproteinase domain 26

Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v1.41

Back Forward Reload Stop Home http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr8:39720414-39814936&hgslid=77522797&refGene=pack&hgFind_match=1 Go

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

UCSC Genome Browser on Human Mar. 2006 Assembly

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

position/search chr8:39,720,414-39,814,936 jump clear size 94,523 bp. configure

Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.

default tracks hide all add custom tracks configure refresh

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

Mapping and Sequencing Tracks

Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb Rate
dense	hide	dense	hide	hide
Map Contigs	Assembly	Gap	Coverage	BAC End Pairs
hide	hide	hide	hide	hide
Fosmid End Pairs	GC Percent	Short Match	Restr Enzymes	
hide	hide	hide	hide	

Genes and Gene Prediction Tracks

Known Genes	RefSeq Genes	Other RefSeq	MGC Genes	Ensembl Genes
pack	pack	hide	pack	hide

Human Gene ADAM2 Description and Page Index

Back Forward Reload Stop Home http://genome.ucsc.edu/cgi-bin/hgGene?hg_gene=BC064547&hg_prot=Q6P2G0_HUMAN **UCSC Known Gene details**

Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help

Human Gene ADAM2 Description and Page Index

Description: ADAM2 protein.
Alternate Gene Symbols: NM_001464
Representative mRNA: BC064547 **Protein:** Q6P2G0
RefSeq Summary: 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.
Position: chr8:39720414-39814886
Strand: -
Genomic Size: 94473
Exon Count: 17 **CDS Exon Count:** 16

[Page Index](#) [Quick Links](#) [Sequence](#) [Microarray](#) [RNA Structure](#) [Protein Structure](#)
[Other Species](#) [GO Annotations](#) [mRNA Descriptions](#) [Methods](#)

Quick Links to Tools and Databases

Genome Browser	Gene Sorter	VisiGene	Proteome Browser	Table Schema	UniProt
Entrez Gene	PubMed	OMIM	GeneLynx	GeneCards	HGNC
CGAP	Stanford SOURCE	ExonPrimer	Ensembl	Jackson Labs	H-INV
Allen Brain Atlas					

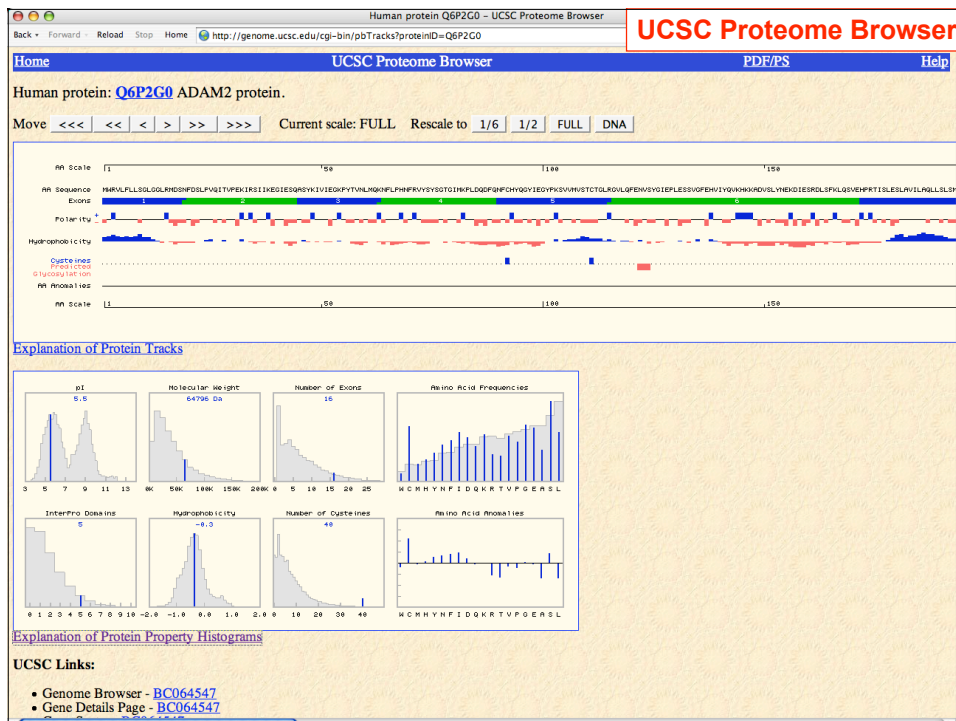
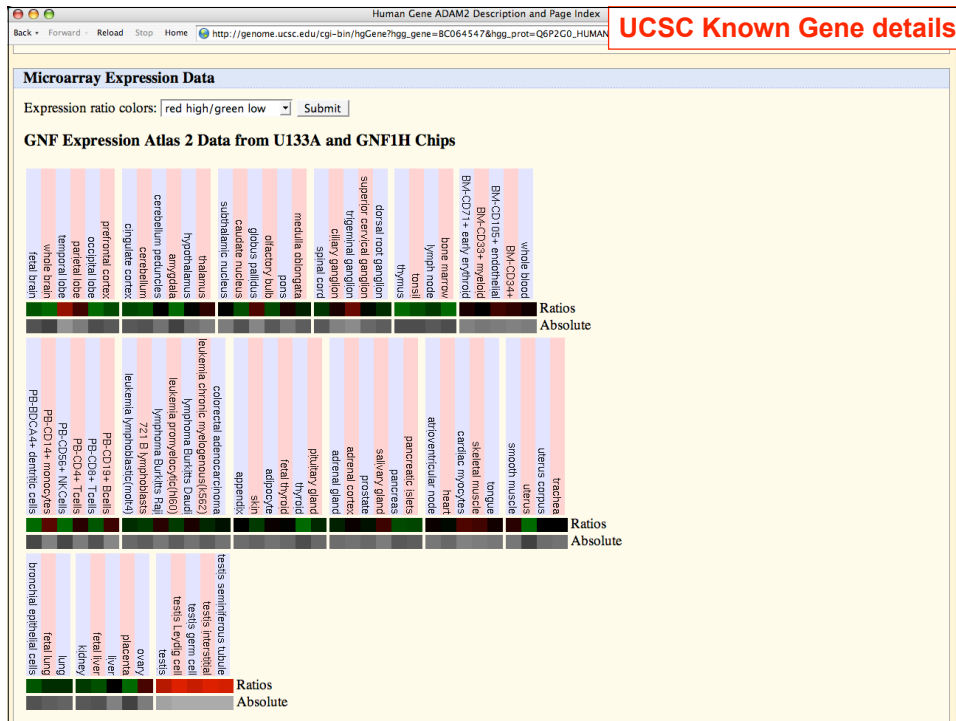
Sequence

Genomic (chr8:39,720,414-39,814,886) mRNA (may differ from genome) Protein (579 aa)

Microarray Expression Data

Expression ratio colors: red high/green low Submit

GNF Expression Atlas 2 Data from U133A and GNF1H Chips



UCSC Genome Browser on Human Mar. 2006 Assembly

position/search chr8:39,720,414-39,814,936 jump clear size 94,523 bp. configure

Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.

default tracks hide all add custom tracks configure refresh

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

Mapping and Sequencing Tracks

Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb Rate
dense	hide	dense	hide	hide
Map Contigs	Assembly	Gap	Coverage	BAC End Pairs
hide	hide	hide	hide	hide
Fosmid End Pairs	GC Percent	Short Match	Restr Enzymes	
hide	hide	hide	hide	

Genes and Gene Prediction Tracks

Known Genes	RefSeq Genes	Other RefSeq	MGC Genes	Ensembl Genes
pack	pack	hide	pack	hide

RefSeq Gene

RefSeq Gene ADAM2

RefSeq: [NM_001464.3](#) Status: Reviewed
 CDS: 3 complete
 OMIM: [601533](#)
 Entrez Gene: [2515](#)
 PubMed on Gene: [ADAM2](#)
 PubMed on Product: [ADAM2 metalloproteinase domain 2 proprotein](#)
 GeneLynx: [ADAM2](#)
 GeneCards: [ADAM2](#)
 AceView: [ADAM2](#)
 Stanford SOURCE: [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.

mRNA/Genomic Alignments

SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
2640	100.0%	8	-	39720414	39814936	NM_001464	1	2640	2657

Position: [chr8:39720414-39814936](#)
 Band: 8p11.22
 Genomic Size: 94523
 Strand: -
 Alternate Name: ADAM2
 CDS Start: complete
 CDS End: complete

Links to sequence:

- [Predicted Protein](#) **click**
- [mRNA Sequence](#) different from the genomic sequence.
- [Genomic Sequence](#) from assembly

UCSC RefSeq Gene details

Genomic Sequence Near Gene

Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help

Get Genomic Sequence Near Gene

Note: if you would prefer to get DNA for more than one feature of this track at a time, try the [Table Browser](#) using the output format sequence.

Sequence Retrieval Region Options:

Promoter/Upstream by bases

5' UTR Exons

CDS Exons

3' UTR Exons

Introns

Downstream by bases

One FASTA record per gene.

One FASTA record per region (exon, intron, etc.) with extra bases upstream (5') and extra downstream (3')

Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Sequence Formatting Options:

Exons in upper case, everything else in lower case.

CDS in upper case, UTR in lower case.

All upper case.

All lower case.

Mask repeats: to lower case to N

1000 nt upstream of ADAM2

```
>hg18_refGene_NM_001464 range=chr8:39814937-39815936
ggaagtatctaccaacacataccctgtgatccgacaactcactctagaa
atatacacagttagaacctcctactatttacoccaaaagcattgagaaga
atgtttatagctaataatttttaaatagctggaacataaacaacaa
aatattcattaacagtaaaatggaacacaagtggttatatttaataga
atttgtatatacaccaatgaggtaaaacagaactatgtcttgataga
accttaacatcactctattaaagaacccagacatgaaagatagttgt
gattgtcttactctcgaaagttcaaaaacagacaaaacgaatcttgg
ttgttgaagtcactggttgaggttggaatctggggatttgggtgggt
ctttttcattttttcaactggtgactagtttaaggtttttttttccac
ttgaatattaatgaactgtgaacttatgattattatataacttttttc
gtttttgttctctctttttttttttttttttttttttttttttttttt
tttgtctctcaccocaggtggagtgcaaggttaaggtttctgtgcac
tgcaacctgctccttaggtcaagcgattctctgctcagcttcccg
agttagtgggatttcaggcaaccgcaacatgctggttaatttttttt
gtatttttagtacagaggggtttcaacatgttgogaactggtctoga
actcctgacgtgtttatataattcaattgaaatttacttaagaagt
gtttataaatcttctgttctcagotgttgaagtgattttgtgttgc
tgttgcttaattaggatcaactccagtggaagtgtctgtctgcagag
acaggtctcaggaagctgcaggttccacagcaccacaccacacactag
cccacctgggtctcccagcgcctacctcttccaggtgctggcggg
```

UCSC

Add tracks to the Genome Browser

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

UCSC Genome Browser on Human May 2004 Assembly

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

position/search chr8:39,720,414-39,814,936 jump clear size 94,523 bp. configure

chr8 (311,22) 39,720,414 39,814,936

UCSC Known Genes (June, 06) Based on UniProt, RefSeq, and GenBank mRNA

RefSeq Genes

RepeatMasker

Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.

move start < 2.0 > move end < 2.0 >

default tracks hide all add custom tracks configure refresh

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

Expression and Regulation

Allen Brain	GNF Atlas 2	GNF Ratio	Affy HnEx 1.0	Affy U133
hide	hide	hide	hide	hide
Affy GNF1H	Affy U133Plus2	Affy U95	CpG Islands	FirstEF
hide	hide	hide	hide	hide
5x Reg Potential	TFBS Conserved	Affy Txn Phase2	SGMO/EIO	NHGRI DNase1-HS
hide	full	hide	hide	hide
Reg Potential 7 species	hide dense squish pack full	PicTar miRNA	hide	hide
hide	hide	hide	hide	hide

Comparative Genomics

Conservation	Most Conserved	Fugu Blat	Fugu Chain	Fugu Net
hide	hide	hide	hide	hide
Tetraodon Ecores	Tetraodon Chain	Tetraodon Net	Zebrafish Chain	Zebrafish Net
hide	hide	hide	hide	hide
X. tropicalis Chain	X. tropicalis Net	Chicken Chain	Chicken Net	Opossum Chain
hide	hide	hide	hide	hide

Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v141

Back Forward Reload Stop Home http://genome.ucsc.edu/cgi-bin/hgTracks?hgid=77541763&hgTrack=77541763&hgTrack2=93EK3E+&position=chr8:39720414-39814936

UCSC TFBS Track

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

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

position/search chr8:39,765,312-39,859,834 jump clear size 94,523 bp. configure

chr8 (311,22) 39,765,312 39,859,834

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

UCSC Known Genes (June, 06) Based on UniProt, RefSeq, and GenBank mRNA

RefSeq Genes

Positional Cloning Gene Collection Full ORF only

Exon (pm) Exon (pm) Human/Mouse/Rest/DOG

Exon (pm) Exon (pm) Human/Mouse/Rest/DOG

Human mRNAs Human ESTs That Have Been Sliced

Sliced ESTs Human Conserved Transcription Factor Binding Sites

Vertebrate Multiple Alignment & Conservation

Conservation

mouse pig rhesus dog chimp human orangutan chimpanzee

SHF Simple Nucleotide Polymorphism (SNP) (not by ESI)

RepeatMasker

Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.

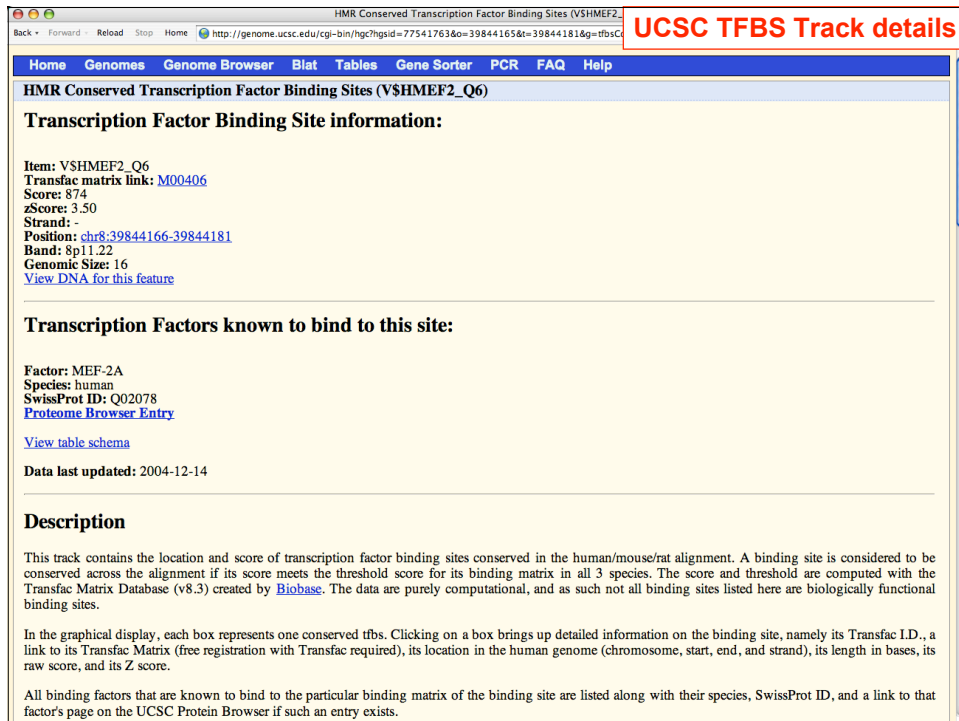
move start < 2.0 > move end < 2.0 >

default tracks hide all add custom tracks configure refresh

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

Mapping and Sequencing Tracks

Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb Rate
dense	hide	dense	hide	hide
Map Contigs	Assembly	Gap	Coverage	BAC End Pairs
hide	hide	hide	hide	hide
Fosmid End Pairs	GC Percent	WSSD Duplication	Short Match	Restr Enzymes



UCSC TFBS Track details

Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help

HMR Conserved Transcription Factor Binding Sites (V\$HMEF2_Q6)

Transcription Factor Binding Site information:

Item: V\$HMEF2_Q6
Transfac matrix link: [M00406](#)
Score: 874
zScore: 3.50
Strand: -
Position: [chr8:39844166-39844181](#)
Band: 8p11.22
Genomic Size: 16
[View DNA for this feature](#)

Transcription Factors known to bind to this site:

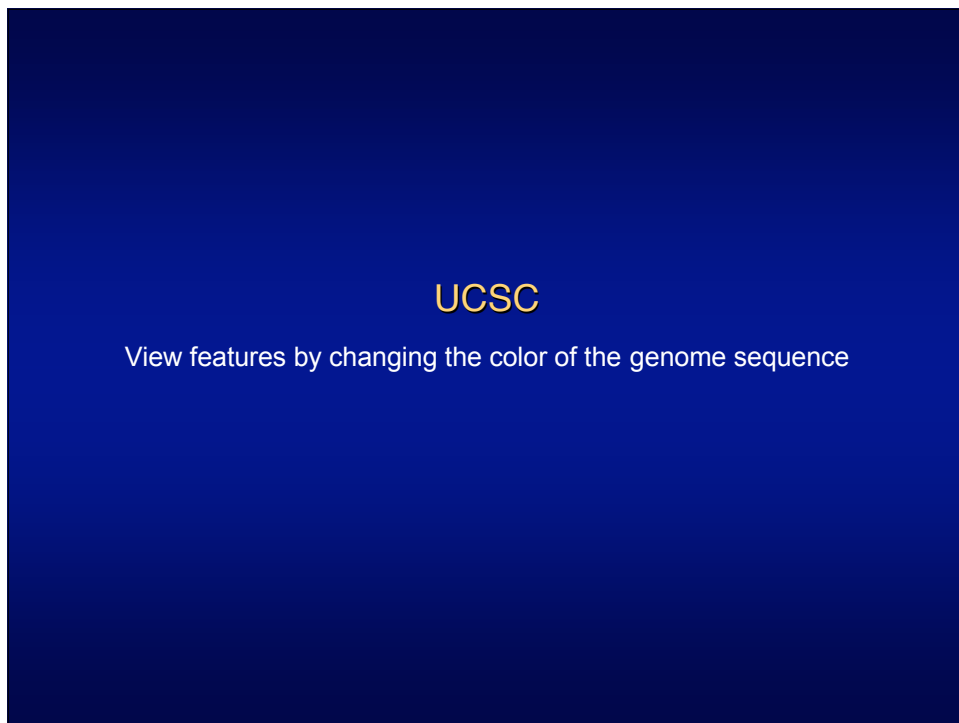
Factor: MEF-2A
Species: human
SwissProt ID: [Q02078](#)
[Proteome Browser Entry](#)
[View table schema](#)
Data last updated: 2004-12-14

Description

This track contains the location and score of transcription factor binding sites conserved in the human/mouse/rat alignment. A binding site is considered to be conserved across the alignment if its score meets the threshold score for its binding matrix in all 3 species. The score and threshold are computed with the Transfac Matrix Database (v8.3) created by [BioBase](#). The data are purely computational, and as such not all binding sites listed here are biologically functional binding sites.

In the graphical display, each box represents one conserved tfbs. Clicking on a box brings up detailed information on the binding site, namely its Transfac I.D., a link to its Transfac Matrix (free registration with Transfac required), its location in the human genome (chromosome, start, end, and strand), its length in bases, its raw score, and its Z score.

All binding factors that are known to bind to the particular binding matrix of the binding site are listed along with their species, SwissProt ID, and a link to that factor's page on the UCSC Protein Browser if such an entry exists.



UCSC

View features by changing the color of the genome sequence



Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v142

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

UCSC Genome Browser on Human Mar. 2006 Assembly

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

position/search chr8:39,720,414-39,814,936 jump clear size 94,523 bp. configure

move start < 2.0 > Click on a feature for details. Click on base position to zoom in around cursor. move end < 2.0 >
 Click on left mini-buttons for track-specific options.

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

Mapping and Sequencing Tracks

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

Variation and Repeats

SNPs	Segmental Dups	RepeatMasker	Simple Repeats	Microsatellite
dense ▾	hide ▾	dense ▾	hide ▾	hide ▾
Self Chain				
hide ▾				

refresh

UCSC SNP Track details

SNPs Track Settings

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

Simple Nucleotide Polymorphisms (dbSNP build 126)

Display mode:

Minimum **Average Heterozygosity**:

Maximum **Weight**:

Any type of data can be excluded from view by deselecting the checkbox below. Not all assemblies include values in all categories.

Location Type:
 Unknown Range Exact Between RangeInsertion RangeSubstitution RangeDeletion

Class:
 Unknown Single Nucleotide Polymorphism In/Del Heterozygous Microsatellite Named No Variation Mixed Mnp Insertion Deletion

Validation:
 Unknown By Cluster By Frequency By Submitter By 2 Hit / 2 Allele By HapMap

Function:
 Unknown Locus Coding - Synonymous Coding - Non-Synonymous Untranslated Intron Splice Site Reference (coding)

Molecule Type:
 Unknown Genomic cDNA

SNP Feature for Color Specification:

The selected feature above has the following values below. For each value, a selection of colors is available.

Unknown: **Locus:** **Coding - Synonymous:** **Coding - Non-Synonymous:** **Untranslated:** **Intron:**
Splice Site: **Reference (coding):**

[View table schema](#)

Description

This track contains dbSNP build 126, available from <ftp.ncbi.nih.gov/snp>.

Interpreting and Configuring the Graphical Display

Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v142

http://genome.ucsc.edu/cgi-bin/hgTracks

Blat Tables Gene Sorter PCR DNA Convert

UCSC Genome Browser on Human Mar. 2006 Assembly

position/search chr8:39,720,414-39,814,936 jump clear size 94,523 bp. configure

UCSC Known Genes based on UniProt, RefSeq, and GenBank mRNA

RefSeq Genes

Human ESTs from GenBank

Human ESTs That Have Been Spliced

Conservation

Simple Nucleotide Polymorphisms (dbSNP build 126)

Red: non-synonymous SNPs
Green: synonymous SNPs
Black: other SNPs



The screenshot displays the NCBI Entrez Protein search interface. The search query is "np_001455". The results show a single entry for "2 proprotein [Homo sapiens] (gi:55743080)". The FASTA sequence is displayed below the entry:

```
>gi|55743080|ref|np_001455.3|ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
MWRVLLSGLGGLRMDNFDLSPVQIVPKRISIRKQIEBQASYKIVIRGKPYVNLQKQKFLPHNF
NVSYSCTKMKPLDQDFNFCYQYIGYPRKSVNVSCTGLRGLQFENVSIGYEDLBSVGFEBVT
YQKRRKADVLELNEQIESLDFPKQVEVQDQYAVYIDNVIYKGLRNGSGDTTVAQKQFELG
LQNAIFVSNITIIISLLELNIDENIAFTGRANELLHFLNKTSTVLVLRHDAFLVIREKSNVGA
TPOGRMDANFAGGVLEPFTIIELELAVLQGLLELSMGITIDDKWQCCGAVCNHPEATHFQVYK
FEMCSFEDPAEFIGKQKQCLINQPLRDFPFKQAVCCNAKLEGBECDCOTQDQCALIGETCCDIATCR
FRAGNCADQPCENCLPMSKRNCRPSFEDCDLPETCGSSASCPENIYVOTGHPCCGLNQVICIDQVCR
SGMKCTPTTKQVEVQSECVHILNKKVYRQKIIISGQYQCCBANDLQCKLTKYVGFELQIPRA
TIIANISGLCIAVEFASDHADQKHWIKDQTSQGNKVCNRQCVSSSLYDCTDKCNDRQVCKNK
KRICGSIYLPFCVQGLMPGSIIDGNFPFVAIPARLERRITENIYHKKRHWPFLLPFFIIFC
VLIAMVKNVQFKKWRTEYSDSDQPSSEPKG
```

UCSC BLAT search

Home Genomes Tables Gene Sorter FAQ Help

Chicken BLAT Search

BLAT Search Genome

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

```
>gi|55743080|zef|NP_001455.3| ADAM metalloproteinase domain 2 proprotein [Homo sapiens]
MNRVFLLSGLRMDGNSFSLPVOITVPEKIRSLIKBIESQASYKIVIRGKPYTVNLQKMFLEHNP
RVYSYSGTGLMKPLDDDFQNFCHYQVIGYPKSVVMVSTCTGLRGLVQFVNSVYGBPLESSVGFHEVI
YQVHKKADVSLYNEKDIERSDLSFKLQSVPEQQDFAKYIEMHVIVKQLYNHMGSDTTVAQRVFLIG
LINAIPVSNITLISLELWIDENKATTOGANELLFTPLRWTSYLVRPHDVAFLLVYREKSNVYGA
TFQCKWQDANYAGGVVHRTFSISRELAIVLQGLLELDMGTYVDLTKCCGSAVCIWVPEAHPISQVYI
FENCSFEDFAHPIISKQSQCLHNQPRLDPPFKQAVCGNAKLEAGEECCDCEBDCALIGTCCDIATCR
FKAGSNCADGPCCECNLFMSKERMCRPSFEBDCDLPYCNSSASCPENHYVQTGHPCGLNQCICIDGVCN
SGDQCTPTFGKEVEFPDSQSYSLSKKTVVSGNCCIISDGYTCCBADNLQCKLTKYVGRFLQIPRA
TLIYANISGHLCAVEFASDHADSQRMKIKDQTSKGNKVCNRQCVSSVLYGDCCTDKCNDRGVCNNK
KHCCHSAYLPDQSVQSLWPGGSDSGNFPVVAIPARLPERRYIENIYHSKPMRWPFPIFFFIIFC
VLIAMVKNVQRRKWRTEYSSDDEQPESESEPKG
```

submit | I'm feeling lucky | clear

Paste 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: Browse... submit file

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

BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 40 bases or more. It may miss more divergent or shorter sequence alignments. It will find perfect sequence matches of 33 bases, and sometimes find them down to 20 bases. BLAT on proteins finds sequences of 80% and greater similarity of length 20 amino acids or more. In practice DNA BLAT works well on primates, and protein blat on land vertebrates.

UCSC BLAT search

Home Genomes Tables Gene Sorter PCR FAQ Help

Chicken BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN	
browser	details	NP_001455.3	44	539	600	735	71.6%	Un	++	6765970	67659720	12
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

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

Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.

default tracks | hide all | add custom tracks | configure | refresh

Chromosome Color Key:
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y M Un

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

Mapping and Sequencing Tracks

Base Position	Supercontigs	Assembly	Gap	BAC End Pairs
dense	hide	hide	hide	hide
GC Percent	Quality Scores	Isochores	Contamination	Short Match
hide	hide	hide	hide	hide

UCSC BLAT search

Chicken BLAT Results

BLAT Search Results

ACTONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	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										

Alignment of NP_001455.3

NP_001455.3
Chicken.chrUn
block1
block2
block3
together

Alignment of NP_001455.3 and chrUn:635370-635555

Click on links in the frame to the left to navigate through the alignment. Matching bases are colored blue and sequence.

NP_001455.3

```

mrvrlflag leglrmdanf dslprvltvp ekirslikog ieeqaaykiv ieeqpytnvl 60
mqknflphnf rvyysagtgi mkpldqdfn fchygyieg ypkavvmvat etqlrgvlf 120
enveyiepl esvgtiehvi yqvkhkadv elynekdies rdlsfkigev epqgfkakyl 180
emvsvveqei ymmsagdtv seqvgtigig llnaifevfn iililaaal widentiaet 240
geanelthf lrwktaylwl rphdvaillv yreksnyva tfgkmodan yaggvvlhpr 300
tisltelevi laqlilsiog ltyddinkoq eegavcimp eaahfsgvki fenesiedia 360
hfisqkqsgc lhnqprldpf ikqavcogna kicageeodc gteqdcailg etocdiator 420
fkagncsaeg pccencifms kemmcrpefe eodlpeycng aaaspenny vtqghpegin 480
qdcidavem seqkqctdfl qkeveopse cyahinaktd vagnocisda vtycccaadn 540
GCKLICKiv gkflilqpra TITVAnisgl L'Lavefaad hadsqkwiK DGTaCCaNV 600
orngrevasa ylgpdeitdk endegvomsk khheasaayl pndcveqadl wpggsidagn 660
fppvaiparl perryienly hokpmrvpff lfipffiiic vilainkvkn lqkkwrted 720
yasdeqpeae sepkg

```

Chicken.chrUn :

```

AATCTGGGCT GTGGAAACT CATCTGcaaa GACccaaac gaggteccct caccacaata 635429
aagggtGCCA TCATCTATGC Tcaagtgcaaa gaacANTCTT Gggtctitt tgaqtaag 635489
catgcaacct ceggsacaga tcctctctgt gttAAGGATG GCACGaaATG CGGTcccgga 635549
AAGGTG

```

Side by Side Alignment*

```

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

001681 T I Y A N I S G H I C 001716
>>>>> | | | | | | | | | | | | | >>>>>
635436 acctcatctatgctcaagtcaagaacatctgtgc 635471

001768 K D G T S C S N K V 001800
>>>>> | | | | | | | | | | | | | >>>>>
635523 aaggatggcaagaaatgggtcccggaagta 635555

```

UCSC

Add your own custom tracks

Human chrX:151,073,054-151,383,976 - UCSC Genome Browser v142

Back Forward Reload Stop Home <http://genome.ucsc.edu/cgi-bin/hgTracks?clade=vertebrate&org=Human&db=hg18&position=chrX3A15192C073K2C054-151X> Go

Add Your Own Custom Track

Display your own custom annotation tracks in the browser using the procedure described in the custom tracks [user's guide](#). For information on upload procedures and supported formats, see the "Loading Custom Annotation Tracks" section below.

Annotation File: Browse... Submit

```

browser position chr22:38496887-39496866
browser hide cytoband
browser hide stsMap
browser hide gap
browser hide clonePos
browser full reGene
browser dense mrna
track name="scale" description="our peak"
chr22 38496887 3896888 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 se146131
chr22 38705963 38705964 se2941443
    
```

Reset Submit

Click [here](#) to view a collection of custom annotation tracks contribute

Nature Genetics: A user's guide to the human genome, Question 7

UCSC Table Browser

- Download track in text format
- Retrieve DNA sequence covered by a track
- Calculate intersections between tracks and view in the Genome Browser. For example:
 - Show all RefSeq genes that contain only one exon
 - Show transcription factor binding sites that overlap (intersect) with a SNP

Table Browser

Use this program to get the data associated with a track in text format, to calculate intersections. See [Using the Table Browser](#) for a description of the controls in this form.

clade: Vertebrate **genome:** Human **assembly:** Mar. 2006
group: Genes and Gene Prediction Tracks **track:** RefSeq Genes
table: refGene [describe table schema](#)
region: genome position chr22:38496887-39496866 [lookup](#)
identifiers (names/accessions): [paste list](#) [upload list](#)
filter: create
intersection: create
correlation: create
output format: all fields from selected table
output file: (leave blank to keep output)
file type returned: plain text gzip compressed
[get output](#) [summary/statistics](#)
 To reset all user cart settings (including custom tracks), [click here](#).

UCSC Table Browser: RefSeq genes that contain only one exon

Filter on Fields from hg18.refGene

bin is ignored

name does match * AND

chrom does match * AND

strand does match * AND

txStart is ignored AND

txEnd is ignored AND

cdsStart is ignored AND

cdsEnd is ignored AND

exonCount is = 1 AND

exonStarts does match *

exonEnds does match *

id is ignored AND

name2 does match * AND

cdsStartStat does match * AND

cdsEndStat does match * AND

exonFrames does match *

AND Free-form query:

[submit](#) [cancel](#)

#	bin	name	chrom	strand	txStart	txEnd	cdsStart	cdsEnd	exonCount	exonStarts	exonEnds	id	name2	cdsStartStat	cdsEndStat	exonFrames
1	#filter: refGene.exonCount = 1															
2	188	NM_006511	chr1	+	15858950	15860803	15858950	15860803	1	15858950	15860803	0	B5C1A1	cmpl	incompl	0
3	178	NM_002232	chr1	-	11015832	11019178	11017226	11018954	1	11015832	11019178	0	KCNM3	cmpl	cmpl	0
4	301	NM_001821	chr1	-	23985790	23986585	23985720	23986581	1	23985790	23986585	0	CHML	cmpl	cmpl	0
5	585	NM_001005484	chr1	+	58953	59871	58953	59871	1	58953	59871	0	OR4F5	cmpl	cmpl	0
6	587	NM_001005221	chr1	+	357521	358458	357521	358458	1	357521	358458	0	OR4F29	cmpl	incompl	0
7	587	NM_001005224	chr1	+	357521	358458	357521	358458	1	357521	358458	0	OR4F3	cmpl	incompl	0
8	587	NM_001005277	chr1	+	357521	358458	357521	358458	1	357521	358458	0	OR4F16	cmpl	incompl	0
9	589	NM_001005221	chr1	-	610960	611897	610960	611897	1	610960	611897	0	OR4F29	incompl	cmpl	0
10	589	NM_001005224	chr1	-	610960	611897	610960	611897	1	610960	611897	0	OR4F3	incompl	cmpl	0
11	589	NM_001005277	chr1	-	610960	611897	610960	611897	1	610960	611897	0	OR4F16	incompl	cmpl	0
12	593	NM_008005	chr1	+	1157507	1160281	1157521	1158511	1	1157507	1160281	0	BIG4L6	cmpl	cmpl	0
13	607	NM_080431	chr1	+	2927905	2929325	2928110	2929244	1	2927905	2929325	0	ACT12	cmpl	cmpl	0

NCBI

View a genomic region between two STS markers

NHGRI Current Topics in Genome Analysis 2006

Mining Genomic Sequence Data

The screenshot shows the NCBI Home Page with a search bar at the top. The main content area is divided into several sections:

- What does NCBI do?**: A section explaining the center's mission, established in 1988, focusing on molecular biology information, public databases, and research in computational biology.
- Hot Spots**: A list of featured resources including Assembly Archive, Clusters of orthologous groups, Coffee Break, Genes & Disease, NCBI Handbook, Electronic PCR, Entrez Home, Entrez Tools, Gene expression omnibus (GEO), Human genome resources, Influenza Virus Resource, **Map Viewer** (highlighted with a red box), dBMHC, Mouse genome resources, My NCBI, ORF finder, Rat genome resources, Reference sequence project, SAGEmap, SKY/CGH database, and dbSNP.
- 100 Gigabases**: A news item celebrating GenBank's milestone of 100 billion bases from over 165,000 organisms.
- PubMed Central**: An archive of life sciences journals, offering free full-text access to over 500,000 articles from over 200 journals.
- NCBI News**: A section for the latest news, including Summer 2006 News available online.

The screenshot shows the NCBI Map Viewer interface. The search bar contains "Homo sapiens (human) Build 36" and "d8s1170 OR d8s94". The results are organized into taxonomic groups:

- Vertebrates**:
 - Mammals**:
 - Bos taurus* (cow)
 - Canis familiaris* (dog)
 - Felis catus* (cat)
 - Homo sapiens* (human) Build 36
 - Homo sapiens* (human) Build 35
 - Macaca mulatta* (rhesus macaque)
 - Mus musculus* (mouse) Build 36
 - Mus musculus* (mouse) Build 35
 - Ovis aries* (sheep)
 - Pan troglodytes* (chimpanzee)
 - Rattus norvegicus* (rat)
 - Sus scrofa* (pig)
 - Other Vertebrates**:
 - Danio rerio* (zebrafish)
 - Gallus gallus* (chicken)
- Invertebrates**:
 - Insects**:
 - Anopheles gambiae* (mosquito)
 - Apis mellifera* (honey bee) Amel_4.0
 - Apis mellifera* (honey bee) Amel_2.0
 - Drosophila melanogaster* (fruit fly)
 - Tribolium castaneum* (red flour beetle)
 - Nematode**:
 - Caenorhabditis elegans* (nematode)
 - Echinoderms**:
 - Strongylocentrotus purpuratus* (purple sea urchin)
 - Protozoa**:
 - Cryptosporidium parvum*
 - Dicystidium discoidium*
 - Pisummodium falcaparum*
- Plants**:
 - Arabidopsis thaliana* (thale cress)
 - Avena sativa* (oat)
 - Beta vulgaris* (beet)
 - Glycine max* (soybean)
 - Hordeum vulgare* (barley)
 - Lotus japonicus* (lotus)
 - Lycopersicon esculentum* (tomato)
 - Manihot esculenta* (cassava)
 - Oryza sativa* (rice)
 - Triticum aestivum* (wheat)
 - Zea mays* (corn)
- Fungi**:
 - Aspergillus fumigatus*
 - Candida glabrata*
 - Cryptococcus neoformans*
 - Debaryomyces hansenii*
 - Encephalitozoon cuniculi*
 - Eremothecium gossypii*
 - Gibberella zeae*
 - Kluyveromyces fragilis*
 - Magnaporthe oryzae*
 - Neurospora crassa*
 - Saccharomyces cerevisiae* (baker's yeast)
 - Schizosaccharomyces pombe* (fission yeast)
 - Ustilago maydis*
 - Yarrowia lipolytica*

At the bottom, there is a section for "See more about" with links for **Bacteria**, **Organelles**, and **Viruses**. A detailed description of the Map Viewer's capabilities is also provided.

NHGRI Current Topics in Genome Analysis 2006
Mining Genomic Sequence Data

Entrez Genome view
http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606&query=d8s1170&200R&20d8s94

NCBI NCBI Map Viewer

Search for d8s1170 OR d8s94 on chromosome(s) assembly All Find

Homo sapiens (human) genome view
Build 36.2 statistics Switch to previous build

Search results for query "d8s1170 OR d8s94": 4 hits

Chr	Assembly	Match	Map	Element	Type	Maps
8	reference	all matches	D8S1170	D8S94	STS	STS
8	Celera	all matches	D8S1170	D8S94	STS	STS

Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

NCBI NCBI Map Viewer

Search Homo sapiens Build 36.2 (Current) BLAST OMM Taxonomy Structure

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

Query: d8s1170 OR d8s94

Master Map: STS
Region Displayed: 32M-44,400K bp

Map 1: Homo sapiens UniGene Clusters
Region Displayed: 32M-44,400K bp
Total Transcript alignments On Chromosome: 0
UniGene Clusters Labeled: 50 Total Transcript alignments in Region: 702
Histogram Data: Tick Width=16,645bp/pixel, Max Height=36 transcripts (logarithmic scale)

Map 2: Genes On Sequence
Region Displayed: 32M-44,400K bp
Total Genes On Chromosome: 984 (1 not localized)
Genes Labeled: 50 Total Genes in Region: 85

Map 3: STS
Region Displayed: 32M-44,400K bp
Total STSs On Chromosome: 817 (11 not localized)
STSs Labeled: 30 Total STSs in Region: 470

NCBI

Change the maps displayed on the Map Viewer

The screenshot displays the NCBI Map Viewer interface. On the left, a sidebar contains navigation links such as 'Search', 'Human genome overview page', and 'Map Viewer Home'. A red arrow labeled 'click' points to the 'Maps & Options' link in the sidebar. The main window shows the 'Maps & Options' dialog box, which is titled 'http://www.ncbi.nlm.nih.gov - Map Viewer'. The dialog includes fields for 'Organism: Homo sapiens', 'Chromosome: 8', and 'Region Shown: 31998795.31-44415851.6.'. It features two panes: 'Available Maps' on the left and 'Maps Displayed (left to right)' on the right. The 'Available Maps' pane lists various map types like 'Sequence Maps', 'Ab Initio', 'Assembly', 'BES Clone', 'Clone', 'Component', 'Contig', 'CpG Island', and 'Ensembl Genes'. The 'Maps Displayed' pane shows 'Phenotype', 'STS', and 'Gene' maps. Below these panes are 'More Options' including 'Show Connections', 'Verbose Mode', 'Compress Map', 'Page Length', and 'Thumbnail View'. The dialog has 'OK', 'Apply', and 'Close' buttons at the bottom.

NHGRI Current Topics in Genome Analysis 2006

Mining Genomic Sequence Data

NCBI Phenotype Map

Map Viewer
<http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?TAXID=9606&CHR=8&EG=31998795.38&END=444158>

Human genome overview page (Build 36.2)
 Human genome overview page (Build 35.1)

Search: **Homo sapiens Build 36.2 (Current)**
 Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
 Query: d8s1170 OR d8s94 [clear](#)

Master Map: Genes On Sequence
 Region Displayed: 32M-44,00K bp

Pheno	STS	Gene_seq	Symbol	O	Links	E	Cyto	Description
			NRG1	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p21-p12	neuregulin 1
			FUT10	+	HGNC sv pr dl ev mm hm	best RefSeq	8p12	fucosyltransferase 10 (alpha (1,3) fucosyltransferase)
			C8orf41	+	HGNC sv pr dl ev mm hm sts	best RefSeq	8p12	chromosome 8 open reading frame 41
			RNF122	+	HGNC sv pr dl ev mm hm	best RefSeq	8p12	ring finger protein 122
			LOC728024	+	sv pr dl ev mm	protein	8p12	similar to CG16865-PA
			RAB11FIP1	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.22	RAB11 family interacting protein 1 (class I)
			ADRB3	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p12-p11.2	adrenergic, beta-3-, receptor
			ASH2L	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.2	ash2 (absent, small, or homeotic)-like (Drosophila)
			STAR	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.2	steroidogenic acute regulator
			LETM2	+	HGNC sv pr dl ev mm hm	best RefSeq	8p12	leucine zipper-EF-hand containing transmembrane protein
			FLJ43582	+	sv pr dl ev mm	best RefSeq	8p12-p11.23	FLJ43582 protein
			TACCC1	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11	transforming, acidic coiled-coil containing protein 1
			HTRA4	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.23	HtrA serine peptidase 4
			ADAM9	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.23	ADAM metalloproteinase domain 9 (meltrin gamma)
			ADAM3A	+	HGNC sv dl ev mm	best RefSeq	8p21-p12	ADAM metalloproteinase domain 3a (cyritestin 1)
			ADAM18	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.22	ADAM metalloproteinase domain 18
			SERP1	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p12-p11.1	secreted frizzled-related protein 1
			GINS4	+	HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.21	GINS complex subunit 4 (Sld5 homolog)
			NKX6-3	+	HGNC sv pr dl ev mm	best RefSeq	8p11.21	NK6 transcription factor related, locus 3 (Drosophila)
			LOC728445	+	sv pr dl ev mm	mRNA	8p11.21	hypothetical protein LOC728445
			PLAT	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p12	plasminogen activator, tissue
			IKBKB	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.2	inhibitor of kappa light polypeptide gene enhancer in B
			DKK4	+	OMIM HGNC sv pr dl ev mm hm	best RefSeq	8p11.2-p11.1	dickkopf homolog 4 (Xenopus laevis)
			CHRNA6	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.21	cholinergic receptor, nicotinic, alpha 6
			THAP1	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.21	THAP domain containing, apoptosis associated protein
			RNF170	+	HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.21	ring finger protein 170
			HOOK3	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.21	hook homolog 3 (Drosophila)

NCBI region between 2 genes

Map Viewer
<http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?TAXID=9606&CHR=8&MAPS=phen>

Human genome overview page (Build 36.2)
 Human genome overview page (Build 35.1)

Search: **Homo sapiens Build 36.2 (Current)**
 Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
 Query: d8s1170 OR d8s94 - not in this view, click [here](#) to see results [clear](#)

Master Map: Genes On Sequence
 Region Displayed: 38,950K-41,290K bp

Pheno	STS	Gene_seq	Symbol	O	Links	E	Cyto	Description
			HTRA4	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.23	HtrA serine peptidase 4
			TM2D2	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.23	TM2 domain containing 2
			ADAM9	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.23	ADAM metalloproteinase domain 9 (meltrin gamma)
			ADAM32	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.23	ADAM metalloproteinase domain 32
			TMDCC1	+	sv pr dl ev mm hm	best RefSeq	8p11.23	TMDCC1
			ADAM3A	+	HGNC sv dl ev mm	best RefSeq	8p21-p12	ADAM metalloproteinase domain 3a (cyritestin 1)
			ADAM18	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.22	ADAM metalloproteinase domain 18
			ADAM2	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.2	ADAM metalloproteinase domain 2 (fertilin beta)
			INDO	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p12-p11	indoleamine-pyrrole 2,3 dioxygenase
			INDOL1	+	HGNC sv pr dl ev mm hm sts	protein	8p11.21	indoleamine-pyrrole 2,3 dioxygenase-like 1
			C8orf4	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.2	chromosome 8 open reading frame 4
			ZMAT4	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.21	zinc finger, matrin type 4
			LOC727725	+	sv pr dl ev mm	best RefSeq	8p11.21	hypothetical protein LOC727725
			SERP1	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p12-p11.1	secreted frizzled-related protein 1



Map Viewer

Back Forward Reload Stop Home <http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?TAXID=9606&CHR=8&MAPS=phen02Csts-rh2Cgenes&QSTR=d8s1170+OR+d8> Go

Homo sapiens Build 36.2 (Current) [BLAST The Human Genome](#)

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

Query: d8s1170 OR d8s94 - **not in this view**, click [here](#) to see results [\[clear\]](#)

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

Region Displayed: 38,950K-41,290K bp [Download/View Sequence/Evidence](#)

Pheno	STS	Gene/seq	Symbol	Q	Links	E	Cyto	Description
38,950K	389500	HTRA4	HTRA4	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.23	HtrA serine peptidase 4
38,950K	389500	TM2D2	TM2D2	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.23	TM2 domain containing 2
38,950K	389500	ADAM9	ADAM9	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.23	ADAM metalloproteinase domain 9 (meltrin gamma)
38,950K	389500	ADAM32	ADAM32	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.23	ADAM metalloproteinase domain 32
38,950K	389500	TMDCH	TMDCH	+	sv pr dl ev mm hm	best RefSeq	8p11.23	tMDC II
38,950K	389500	ADAM3A	ADAM3A	+	HGNC sv dl ev mm	best RefSeq	8p21-p12	ADAM metalloproteinase domain 3a (cyttestin 1)
38,950K	389500	ADAM18	ADAM18	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.22	ADAM metalloproteinase domain 18
38,950K	389500	ADAM2	ADAM2	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.2	ADAM metalloproteinase domain 2 (fertilin beta)
38,950K	389500	INDO	INDO	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p12-p11	indoleamine-pyrrole 2,3 dioxygenase
38,950K	389500	INDOL1	INDOL1	+	HGNC sv pr dl ev mm hm sts	protein	8p11.21	indoleamine-pyrrole 2,3 dioxygenase-like 1
38,950K	389500	C8orf4	C8orf4	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.2	chromosome 8 open reading frame 4
38,950K	389500	ZMAT4	ZMAT4	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.21	zinc finger, matrin type 4
38,950K	389500	LOC727725	LOC727725	+	sv pr dl ev mm	best RefSeq	8p11.21	hypothetical protein LOC727725
38,950K	389500	SFRP1	SFRP1	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p12-p11.1	secreted frizzled-related protein 1

Done

Gene
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=full_report&list_uids=2515

Entrez Gene

NCBI
Search Gene for [] Go Clear

Display Full Report Show 5 Send to

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

1: ADAM2 ADAM metallopeptidase domain 2 (fertilin beta) [Homo sapiens]
GeneID: 2515 Primary source: HGNC:198 updated 17-Sep-2006

Summary

Official Symbol: ADAM2 and **Name:** ADAM metallopeptidase domain 2 (fertilin beta) provided by HUGO Gene Nomenclature Committee
See related: HPRD:03322, MIM:601533
Gene type: protein coding
Gene name: ADAM2
Gene description: ADAM metallopeptidase domain 2 (fertilin beta)
RefSeq status: Reviewed
Organism: *Homo sapiens*
Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Gene aliases: FTNB; PH30; CRYN1; CRYN2; PH-30b
Summary: 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.

Genomic regions, transcripts, and products

(minus strand) RefSeq below

Genomic context See ADAM2 in MapViewer

chromosome: 8; Location: 8p11.2

Bibliography Gene References into Function (GeneRIF): Submit

PubMed links

Interactions

Description

ADAM2 Product	Interactant	Other Gene	Complex	Source	Pubs

Gene
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=full_report&list_uids=2515

Entrez Gene

General gene information

Markers
RH70674(c-PCR) (Links: UniSTS:33582)
Alternate name: U52370
SHGC-111659(c-PCR) (Links: UniSTS:168466)
ADAM2_2713(c-PCR) (Links: UniSTS:462012)

GeneOntology Provided by GOA

Function	Evidence
integrin binding	TAS PubMed
metalloendopeptidase activity	IEA
protein binding	IEA
Process	
cell adhesion	IEA
fusion of sperm to egg plasma membrane	TAS PubMed
proteolysis	IEA
Component	
integral to plasma membrane	TAS PubMed
membrane	IEA

Homology: Mouse, Rat
Map Viewer

General protein information

Names: ADAM metallopeptidase domain 2
fertilin beta; a disintegrin and metalloproteinase domain 2; a disintegrin and metalloproteinase domain 2 (fertilin beta)

NCBI Reference Sequences (RefSeq)

mRNA Sequence NM_001464
Source Sequence BG719616, B1460477, U38805
Product NP_001455 ADAM metallopeptidase domain 2 proprotein

Conserved Domains (4) summary

- pfam01421: Reprolysin; Reprolysin (M12B) family zinc metalloprotease
Location: 178 - 375 Blast Score: 626
- pfam01562: Pep_M12B_prosp; Reprolysin family propeptide
Location: 62 - 174 Blast Score: 362
- smart00050: DJISIN; Homologues of snake disintegrins ; Snake disintegrins inhibit the binding of ligands to integrin receptors
Location: 393 - 470 Blast Score: 281
- smart00608: ACR; ADAM Cysteine-Rich Domain
Location: 472 - 609 Blast Score: 375

Related Sequences

Nucleotide	Protein
Genomic AC136365	None
Genomic AP005902 (96106..113980, complement)	None
mRNA A1133005	CAB40813
mRNA BC024957	AAH34957
mRNA BC064547	AAH64547
mRNA BG719616	None
mRNA B1460477	None

OMIM - A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 2; ADAM2

http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=601533

OMIM

NCBI

MIM #601533
Description
Cloning
Gene Function
Mapping
Animal Model
History
References
Contributors
Citation Data
Edit History

Gene map

Ensembl Gene
Nomenclature
RefSeq
GenBank
Protein
UniGene
LinkOut

Search: OMIM for

Limits Preview/Index History Clipboard Details

Display: Detailed Show 20 Send to

All: 1 OMIM dBSNP: 0 OMIM UniSTS: 0

***601533**
A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 2; ADAM2

Alternative titles; symbols
FERTILIN, BETA; FTNB
PH30

Gene map locus 8p11.2

TEXT

DESCRIPTION

The ADAMs (a disintegrin and metalloprotease domain) are a family of type I transmembrane glycoproteins that share homology with snake venom metalloprotease/disintegrins and sperm surface proteins. They are important in diverse biologic processes such as cell adhesion and proteolytic shedding of cell surface receptors. Structurally, ADAMs consist of a prodomain that blocks protease activity; a zinc-binding metalloprotease domain; disintegrin and cysteine-rich domains with adhesion activity; an epidermal growth factor (EGF [131530]-like domain with cell fusion activity; a transmembrane domain; and a phosphorylated cytoplasmic regulatory domain. For a review of the ADAM gene family, see Primakoff and Myles (2000).

CLONING

Gupta et al. (1996) cloned human fertilin-beta (ADAM2). Using a guinea pig fertilin-beta cDNA as a probe to screen a human testis cDNA library, they obtained the 5-prime end of the human fertilin-beta cDNA using RACE PCR. The complete human fertilin-beta cDNA contains an open reading frame of 2,199 bp and a 380-bp 3-prime untranslated region. Gupta et al. (1996) compared the sequence of human fertilin-beta to related proteins and found metalloprotease, disintegrin, cysteine-rich, EGF-like repeat and transmembrane domains, a structural organization consistent with other members of the metalloprotease/disintegrin family. The amino acid sequence of the mature human fertilin-beta is 90% identical to monkey fertilin and 56 to 59% identical to mature mouse and guinea pig fertilin-betas. Gupta et al. (1996) performed Northern blot analysis of human fertilin-beta which detected a 3.2-kb transcript only in testis RNA, suggesting that human fertilin-beta may be specific to the testis.

Independently, Burkin et al. (1997) cloned a human FTNB cDNA. They stated that the cDNA encodes a predicted 735-amino acid precursor protein from which the signal sequence (amino acids 1 to 16) and metalloprotease domain (amino acids 17 to 382) are cleaved during maturation.

GENE FUNCTION

Gupta et al. (1996) stated that most snake venom disintegrins contain the consensus integrin-binding sequence RGD. They noted that guinea pig, mouse, monkey, and human fertilin-betas contain tripeptide sequences TDE, QDE, FDE, and FEE at this location, respectively. These tripeptides are believed to mediate interaction with an integrin on the surface of the egg and thus mediate sperm/egg binding.

HomoloGene

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=homologene&dopt=HomoloGene&list_uids=1127

HomoloGene (hm)

NCBI

Discover Homologs

My NCBI [Sign In] [Register]

Search: HomoloGene for

Limits Preview/Index History Clipboard Details

Display: HomoloGene Show 20 Send to

All: 1 Fungi: 0 Mammals: 1

I: HomoloGene:1127. Gene conserved in Eutheria

Download, Links

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

H.sapiens ADAM2
ADAM metalloprotease domain 2 (fertilin beta)

P.troglodytes ADAM2
ADAM metalloprotease domain 2

C.familiaris LOC79576
similar to ADAM 2 precursor (A disintegrin and metalloprotease domain 2) (Fertilin beta subunit) (PH30) (PH30)

M.musculus Adam2
a disintegrin and metalloprotease domain 2

R.norvegicus Adam2
a disintegrin and metalloprotease domain 2

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

NP_001455.3
735 aa

XP_519722.1
679 aa

XP_532795.2
881 aa

NP_033748.1
735 aa

NP_064462.1
739 aa

Alignment Scores
Various evolutionary parameters derived from pairwise alignments have been saved.

Show Table of Pairwise Scores

Alignments can be regenerated using BLAST for any selected pair of proteins.

Regenerate Alignments

NP_001455.3(H.sapiens, ADAM2)

XP_519722.1(P.troglodytes, ADAM2)

BLAST

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

smart00050

DISIN, DISIN

smart00058

ACR, ACR

pfam01562

Pep_M12B_propep, Pep_M12B_propep

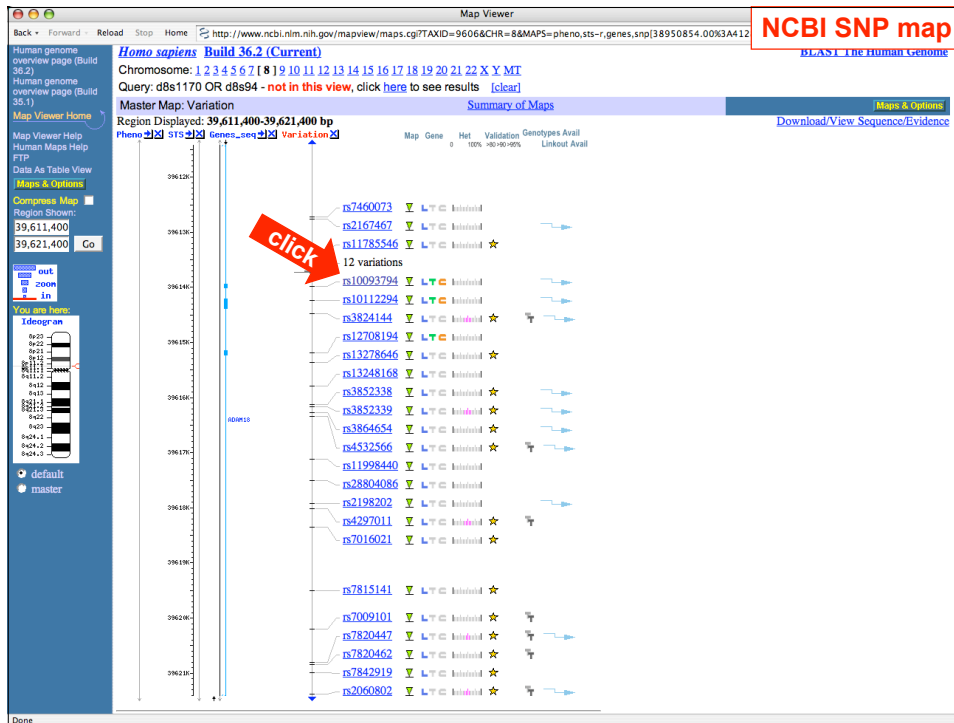
pfam1421

Reprolysin, Reprolysin

Related Homology Resources
Links to curated and computed homology information found in other databases.

MGI:1340894
Orthology group for M.musculus Adam2 includes H.sapiens ADAM2 and R.norvegicus Adam2.

Phenotypes
Phenotypic information for the genes in this entry imported from model organism databases.



The screenshot shows the NCBI dbSNP interface for rs10093794. The top navigation bar includes "PubMed", "Nucleotide", "Protein", "Genome", "Structure", "PopSet", "Taxonomy", "OMIM", and "Books SNP". The main content area displays the following information:

- rs10093794** Allele Links
- Organism: human (*Homo sapiens*)
- Molecule Type: Genomic
- Created/Updated in build: 119/126
- Map to Genome Build: 36.1
- Variation Class: SNP: single nucleotide polymorphism
- Alleles: G/T
- Ancestral Allele: G

SNP Details are organized in the following sections: [Submission](#) | [Fasta](#) | [Resource](#) | [GeneView](#) | [Map](#) | [Diversity](#) | [Validation](#) | [Linkout](#)

Fasta sequence (Legend)
>gnldbSNPrs10093794:allelePos=501:totalLen=701:taxid=9606:snpclass=1:alleles=G/T:mol=Genomic:build=126

```

tgaacaccac cttttcaact tgtttcaatg tattggaata tattgtttt gttttgttta
gataccacat ataagtlaga toatgtagta tttttcttct tgaatoggo ttatttcaat
tagcaaatgt tctctcaatg tctctcaatg tttcaacaaat gaaacagaaa ccaatttcaa
ggcttaataa tatttccgaa tgtgtgat? Ttatatatat atatatatat atatatatat
atatatatat atatatatgt atatacaaaa tatataTCC ATAAAAATCT? GTtcacataa
aaatattgta gaatttgtct aaattgtgac acattctctc atgtacttita aaacagctct
agattacttla taatatctaa tacaatgoc? aaatatact? teatlocat? ggatcaaaaa
taatactTCC ATGGAATAA? AAATTTCTTT? TCCGATGGA? TTATATGGA? TCTGAAATGA
TCCGCTAAC ACAAAAAAT?
&
TCCAGGTTAT TGGGCTGTC AACACTGTAA GTTTTACTT TTTACACTT CCATTTTCAT
GAAAGTTTCT TTAATAAAT? TFGGTTCTCA TCTTGCCAA TGAATAACT? AAATACCTTT
TATTGTTTCA ATCTGAAAT? TACTGGAAAC TTTTCTTAT? TTTTATAGT ACCTTTTAA
TGTAATAACT TTTTCTTAA
    
```

GeneView
GeneView via analysis of contig annotation: ADAM18 a disintegrin and metalloproteinase domain 18
Click to see [\[all\]](#) [\[cSNP\]](#) [\[has frequency\]](#) [\[double hit\]](#) [\[haplotype tagged\]](#) variations associated with this gene.

Group Label	Contig->mRNA	Gene Model (contig mRNA transcript)	Color Legend
reference	NT_007995->NM_014237	sv function	
Celera	NW_923907->NM_014237	sv function	

Group label	Contig->mRNA->Protein	Contig position	mRNA orientation	mRNA pos	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos
reference	NT_007995->NM_014237->NP_058052	9815217	forward	634	nonsynonymous	T	Phe [F]	1	212
					contig reference	G	Val [V]	1	212
Celera	NW_923907->NM_014237->NP_058052	2701448?	forward	634	nonsynonymous	T	Phe [F]	1	212
					contig reference	G	Val [V]	1	212



BLAST Chicken Sequences.

Back Forward Reload Stop Home <http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9031> **NCBI BLAST search**

NCBI Home > Genomic Biology > Chicken Genome Resources > BLAST

Search | Map Viewer | Clear

BLAST
Overview
FAQs
News
Manual
References
Retrieve results
Genome Project

BLAST Chicken Sequences.

Enter an accession, gi, or a sequence in FASTA format:
sp_001455

Or, choose a file to upload
Browse...

Set subsequence: (optional)
From: To:

Database:
genome (reference only) 40216 sequences

Program:
tblastn: Compare a protein sequence against a nucleotide database **select**

Optional parameters

Expect	Filter	Descriptions	Alignments
0.01	low complexity	100	100

Advanced options:

Begin Search Clear Input

Get the URL with preset values? Get URL

Genome View Show positions of the BLAST hits in the chicken genome using the Entrez Genomes MapViewer

NCBI BLAST search

Query= gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2 propeptin [Show maplet] Length=735

Distribution of 40 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

Color key for alignment scores

Sequences producing significant alignments:	Score (Bits)	E Value
ref NW_096640.1 GgaUn_WGA36448.1 Gallus gallus chromosome Un ...	359	1e-97
ref NW_060531.1 Gga15_WGA344.1 Gallus gallus chromosome 15 ge...	189	2e-46
ref NW_060143.1 GgaUn_WGA673.1 Gallus gallus chromosome Un ge...	53.2	4e-18
ref NW_060339.1 Gga6_WGA75.1 Gallus gallus chromosome 6 geno...	23.5	1e-10
ref NW_060253.1 Gga1_WGA62.1 Gallus gallus chromosome 1 genom...	68.6	5e-10

```

>ref|NW_096640.1|GgaUn_WGA36448.1| Gallus gallus chromosome Un genomic contig, whole genome shotgun
sequence
Length=35588

Score = 359 bits (922), Expect = 1e-97
Identities = 215/681 (31%), Positives = 345/681 (50%), Gaps = 36/681 (5%)
Frame = +1

Query 4      VLFLLSGLQLRMDNFDLSPVQIT----VPEKIRSIKEIGESGAEYKIVIECKPYTV 58
             VL +L GL G + +S P+ +T VP ++ S + + SV + +ECFP +
Sbjct 19795  VLVVLLGLVGCPTTDDDESPLVTCMNVTPROL-SPRADFNPLTVSYNQLVQGRQVL 19971

Query 59      NLM-QKNFLPHNFRVSYSGTCKMKPLDQDFNCFHYQYIEGYPKSVVMVSTC-TCLRG 116
             L +K F + +Y G + O+ C YG +G P S+V + TC GLRG
Sbjct 19972  RLRFKGLASRPFITVYDEGAAARBEQVYVDMCFITGCEVGSSEGLVALGTCGCLRG 20151

Query 117     VLOFNVSYGIEPIESSVGFPHYIYQVHKKADVS----LYNEKDIERSDLFKLQVPEP 172
             VL E +Y IEP+ F+H++Y+++ + L E+ + + LQ+
Sbjct 20152  VLMNBSGYEIEPIPDPAFQHLRYMEADSDMPGPTCGLTFEELQYQKTVLPWLPQAPRT 20331

Query 173     QQDF-----AKYIEMHVIEKQLNMGSDTTVVAQKVFQGLIGLTAIIVSFNITII 224
             + + + + + + + + + + + + + + + + + + + + + + + + + + + + + + +
Sbjct 20332  EDKYM LKDNWHTRYKLVVVDNVRFRVRSRDNESKVLROVLEVVNICDLSYDQLSGLVF 20511

Query 225     LSSLELNIDENKIAATGEANELLHTFLRNKTSYLVLNR-PHDVAFLLVYREKSNVYGATFO 283
             L LE+W + N I T A++ L F RN+ S L R HD A L ++ +C +
Sbjct 20512  LVCLSIWVSNPVIWTKSASKTLADFNWRKSDLSLVRMQUDTFLIPAFQCFKSLGLAYL 20691
    
```

Gallus gallus (chicken) genome view
Build 1.1 statistics

Color key for scores: <40 (black), 40-50 (blue), 50-80 (green), 80-200 (magenta), >=200 (red)

NCBI BLAST search

Gallus gallus (chicken) Build 1.1

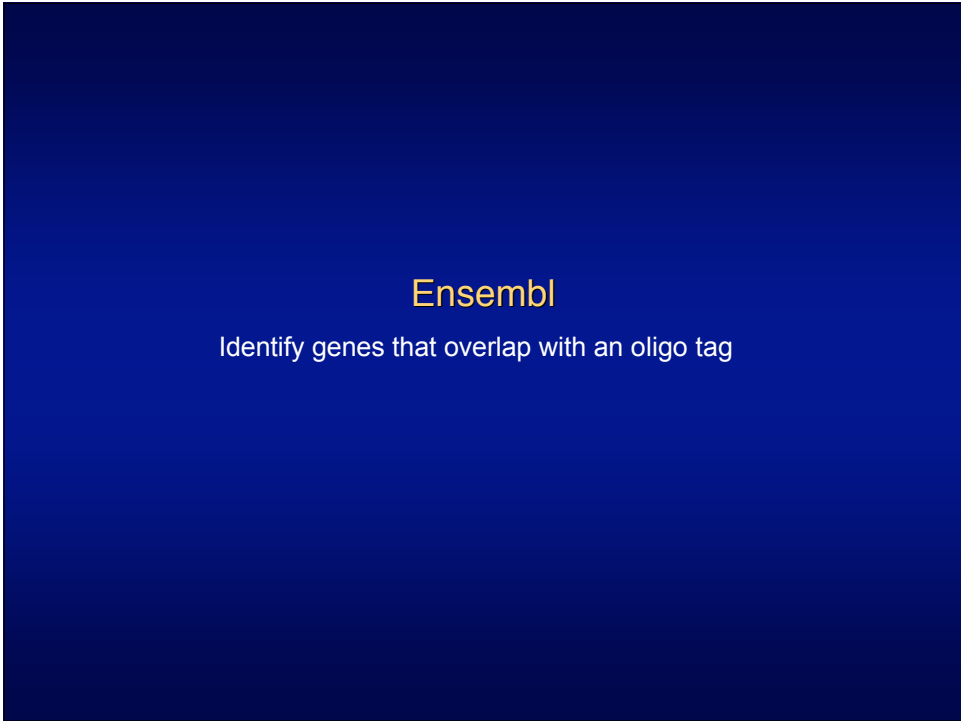
Chromosome: Unknown
Contig: NW_096640 (not localized on the chromosome)

Query: BLAST: gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2

Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200

Master Map: Contig
Region Displayed: 19,550-22,050 bp

Hit: NW_096640.1 ? Identity=31% 662..4



Ensembl Genome Browser
http://www.ensembl.org/index.html

Ensembl release 40 - Aug 2006

Use Ensembl to...
Run a BLAST search
Search Ensembl
Data mining (BioMart)
Export data
Download data

Docs and downloads
Information
What's New
About Ensembl
Ensembl data
Software

Other links
Home
Sitemap
Vega
Pre Ensembl
View previous release of page in Archival
Stable Archival link for this page
Archival sites
Trace server

What's New in Ensembl 40
click
New low-coverage genomes (*L. africana*, *novemcinctus*, *E. telfairi*, *O. curvicolus*)
Sickleback assembly and genebuild (*Gasterosteus aculeatus*)
New species - *Aedes aegypti* (*Aedes aegypti*)
New Macaque assembly and genebuild (*Macaca mulatta*)
New genebuild on Rat assembly (*Rattus norvegicus*)
More news...
About Ensembl
Ensembl is a joint project between EMBL, EBI and the Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes. Ensembl is primarily funded by the Wellcome Trust.
This site provides free access to all the data and software from the Ensembl project. Click on a species name to browse the data.
Access to all the data produced by the project, and to the software used to analyse and present it, is provided free and without constraints. Some data and software may be subject to third-party constraints.
For all enquiries, please contact the Ensembl HelpDesk (helpdesk@ensembl.org).

Other sites using the Ensembl system
EBI Genome Reviews database - mainly archaea and bacteria.
VEGA - Vertebrate Genome Annotation
More...

Mammalian genomes
Homo sapiens (NCBI 36 | Vega)
Pan troglodytes (PanTro 1.0 | **NEW!** *pre!*)
Macaca mulatta (UPDATED) (PPUL 1.0)
Mus musculus (NCBI m36 | Vega)
Rattus norvegicus (UPDATED) (RSC 3.4)
Oryctolagus cuniculus (**NEW!** RABBIT)
Canis familiaris (Canfam 1.0 | Vega | **UPDATED!** *pre!*)
Bos taurus (Btau 2.0)
Sus scrofa (**NEW!** (clone status map)
Dasyurus novemcinctus (**NEW!** AP1A)
Loxodonta africana (**NEW!** BROAD E1)
Echinops telfairi (**NEW!** TENREC)
Monodelphis domestica (MonDom 4)
Ornithorhynchus anatinus (**Pre!** **NEW!** OANA 5)

Other species
Gallus gallus (VASHUC 1)
Xenopus tropicalis (Xt1 4.1)
Danio rerio (Zv6 | Vega)
Takifugu rubripes (FUGU 4.0)
Tetraodon nigroviridis (TETRAODON 7)
Gasterosteus aculeatus (**NEW!** BROAD 51)
Oryzias latipes (MEDAK 1)
Ciona intestinalis (JGI2)
Ciona savignyi (CSAV 2.0)
Drosophila melanogaster (UPDATED) (EGGP 4)
Anopheles gambiae (AganP3)
Aedes aegypti (**NEW!** Aaag1.1)
Caenorhabditis elegans (VS 150)
Saccharomyces cerevisiae (SBD 1)

© 2006 WTS / EBL. Ensembl is available to download for public use - please see the code licence for details.

NHGRI Current Topics in Genome Analysis 2006

Mining Genomic Sequence Data

Ensembl BLAST search

Ensembl release 40 - Aug 2006

Use Ensembl to...

- Run a BLAST search
- Search Ensembl
- Data mining [BioMart]
- Export data
- Download data

Docs and downloads

- Information
- What's New
- About Ensembl
- Ensembl data
- Software

Other links

- Home
- Stemap
- Vega
- Pre Ensembl
- View previous release of page in Archivel
- Stable Archivel link for this page
- Archivel sites
- Trace server

Enter the Query Sequence

Either Paste sequences (max 30 sequences) in FASTA or plain text:

```
MPSS_1
AAAAAATGCCCCCTGAGAG
```

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:

Select the databases to search against

Select species: Use 'ctrl' key to select multiple species

Genomic sequence

Select the Search Tool

BLASTN

Search sensitivity: Near-exact matches (oligo)

About BlastView

© 2006 WTS / EBI. Ensembl is available to download for public use - please see the code licence for details.

Ensembl BLAST search

Alignment Locations vs. Karyotype (click arrow to hide)

Alignment Locations vs. Query (click arrow to hide)

Alignment Summary (click arrow to hide)

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

Query	Subject	Chromosome	Start	End	ori	Score	E-val	SIID	Length
[A] [S] [G] [C]	1	Chr:15	54998168	54998187	+	20	0.0063	100.00	20
[A] [S] [G] [C]	1	Chr:8	72205113	72205129	-	17	0.39	100.00	17

© 2006 WTS / EBI. Ensembl is available to download for public use - please see the code licence for details.

100% identity over 100% of the query length

click

NHGRI Current Topics in Genome Analysis 2006
Mining Genomic Sequence Data

Ensembl v40: Homo sapiens Features on Chromosome 15 54996168-55000187

Ensembl Human ContigView

Chromosome 15 54,996,168 - 55,000,187

Overview

Chr. 15 band

DNA(contigs)

Markers

Ensembl Genes

nRNA Genes

EST Genes

Gene legend

© 2006 WTS / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

Ensembl v40: Homo sapiens Features on Chromosome 15 54996168-55000187

Ensembl Human ContigView

Chromosome 15 54,996,168 - 55,000,187

Overview

Detailed view

Features

Jump to region 15

Chr. 15 Length

EMBL mRNAs

Unigene

GenScan

EST trans

Ensembl trans

Blast hits

DNA(contigs)

Gene legend

© 2006 WTS / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

The screenshot displays the Ensembl v40 ContigView interface for Homo sapiens features on Chromosome 15 (region 54996168-55000187). The browser address bar shows the URL: http://www.ensembl.org/Homo_sapiens/contigview?panel_bottom=off;w=15%3A54996168-55000187;h=BL. A red box highlights the text "Ensembl ContigView" in the top right corner.

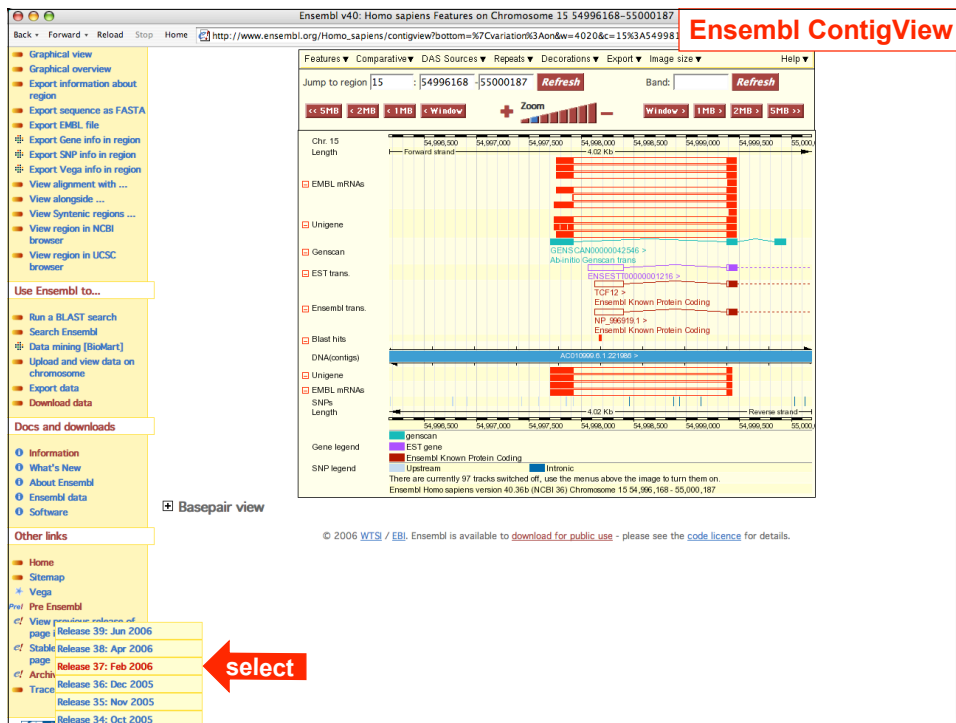
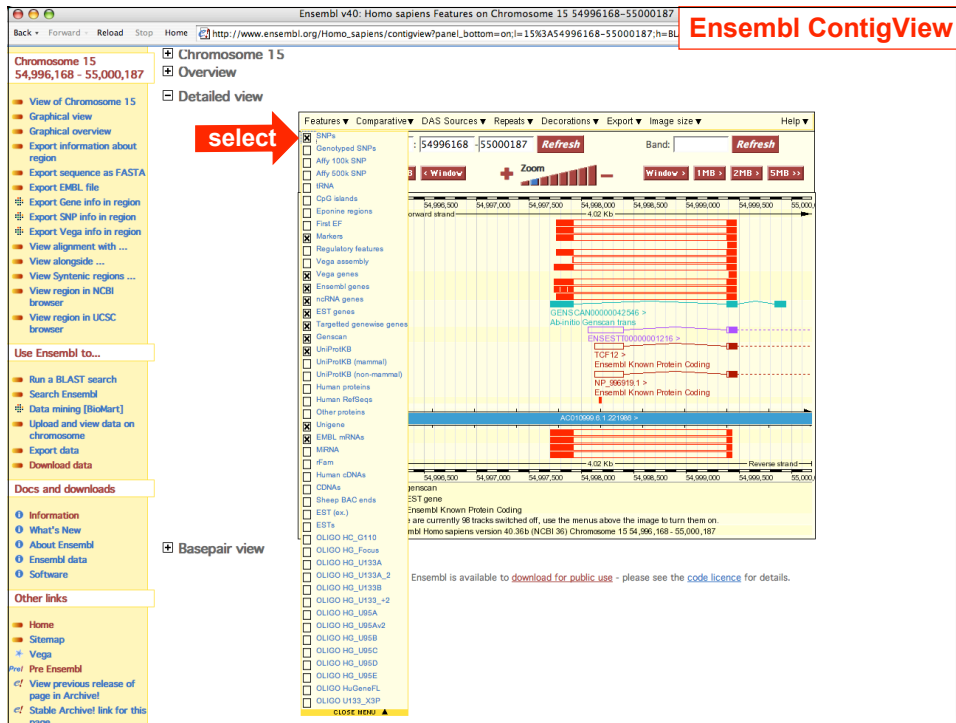
The interface is divided into a left sidebar and a main content area. The sidebar contains several sections:

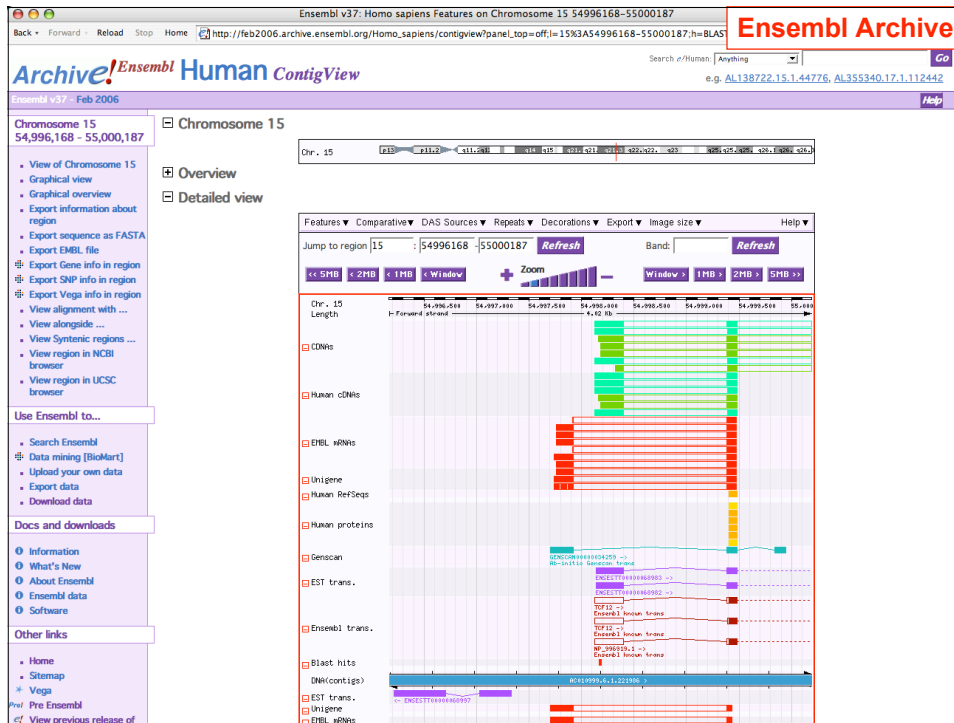
- Graphical view**: Graphical overview, Export information about region, Export sequence as FASTA, Export EMBL file, Export Gene info in region, Export SNP info in region, Export Vega info in region, View alignment with ..., View alongside ..., View Systemic regions ..., View region in NCBI browser, View region in UCSC browser.
- Use Ensembl to...**: Run a BLAST search, Search Ensembl, Data mining [BioMart], Upload and view data on chromosome, Export data, Download data.
- Docs and downloads**: Information, What's New, About Ensembl, Ensembl data, Software.
- Other links**: Home, Stemap, Vega, Pre Ensembl, View previous release of page in Archive!, Stable Archive! link for this page, Archive! sites, Trace server.

The main content area is titled "Basepair view" and shows a genomic track for Chr. 15. The track includes:

- Length**: 100 bp scale from 54,996,150 to 54,996,250.
- Genscan**: GENS CAN000004256 > Ab-initio Genscan frame.
- EST trans.**: ENSEST700000012 > YCF 12 > Ensembl Known Protein Coding.
- Ensembl trans.**: NP_56616.1 > Ensembl Known Protein Coding.
- Blast hits**: A list of protein sequences with their amino acid translations, such as Sth1321, Aha1, Boc2, Bms5, Bosh, Ecol1, Aha1, Ahy1, Ama8, Bms1, Bsa11, Bsk4, Csf44, Ecol16, and N338.

Ensembl
Add features to the ContigView





Ensembl

Get additional information about the gene, transcripts, and exons

Ensembl ContigView

Ensembl v40: Homo sapiens Features on Chromosome 15 54996168-55000187

Search of Human: [Anything] **Go**

e.g. AL138722.15.1.44776, AL355340.17.1.112442

Ensembl release 40 - Aug 2006

Chromosome 15
54,996,168 - 55,000,187

Overview
Detailed view

View of Chromosome 15
Graphical overview
Export information about region
Export sequence as FASTA
Export EMBL file
Export Gene info in region
Export SNP info in region
Export Vega info in region
View alignment with ...
View alongside ...
View Syntenic regions ...
View region in NCBI browser
View region in UCSC browser

Use Ensembl to...

Run a BLAST search
Search Ensembl
Data mining [BioMart]
Upload and view data on chromosome
Export data
Download data

Docs and downloads

Information
What's New
About Ensembl
Ensembl data
Software

Other links
Home
Stemap
Vega

Features ▾ Comparative ▾ DAS Sources ▾ Repeats ▾ Decorations ▾ Export ▾ Image size ▾ Help ▾

Jump to region 15 : 54996168 - 55000187 Refresh Band: Refresh

Chr 15
Length
Forward strand

EMBL mRNAs
Unigene
GenScan
EST trans.
Ensembl trans.
Blast hits
DNA(contigs)
Unigene
EMBL mRNAs
SNPs

Gene legend
EST gene
Ensembl Known Protein Coding
Intronic
Upstream

SNP legend
Upstream
Intronic

There are currently 97 tracks switched off, use the menus above the image to turn them on.
Ensembl Homo sapiens version 40.36b (NCBI 36) Chromosome 15 54,996,168 - 55,000,187

© 2006 WTS / EBI. Ensembl is available to download for public use - please see the code licence for details.

Ensembl GeneView

Ensembl v40: Homo sapiens Gene report for ENSG00000140262

Search of Human: [Anything] **Go**

e.g. ENSG00000139618, ENSG00000128573

Ensembl release 40 - Aug 2006

ENSG00000140262

Gene information
Gene splice site image
Gene regulation info.
Genomic sequence
Genomic sequence alignment
Gene tree info
Gene variation info.
ID history
Transcript information
Exon information
Protein information
Export gene data

Chromosome 15
54,998,125 - 55,368,004

View of Chromosome 15
Graphical overview
Export information about region
Export sequence as FASTA
Export EMBL file
Export Gene info in region
Export SNP info in region
Export Vega info in region

Use Ensembl to...

Run a BLAST search
Search Ensembl
Data mining [BioMart]
Upload and view data on chromosome
Export data
Download data

Docs and downloads

Information
What's New

Ensembl Gene Report for ENSG00000140262

Gene
TCF12 (HGNC Symbol ID) - To view all Ensembl genes linked to the name [click here](#).
This gene is a member of the human CCDS set: [CCDS10159](#), [CCDS10160](#)

Ensembl Gene ID
ENSG00000140262

Genomic Location
This gene can be found on Chromosome 15 at location [54,998,125-55,368,004](#).
The start of this gene is located in [Contig AC010999.6.1.221986](#).

Description
Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SwissProt Q99061](#)

Prediction Method
Genes were annotated by the Ensembl automatic analysis pipeline using either a GeneWise/Exonerate model from a database protein or a set of aligned cDNAs followed by an ORF prediction. GeneWise/Exonerate models are further combined with available aligned cDNAs to annotate UTRs (For more information see V.Curwen et al., Genome Res. 2004 14:942-50.)

Transcripts

Transcript ID	Start	End	Transcript info	Exon info	Peptide info
ENST00000267811	547811	TCF12	[Transcript info]	[Exon info]	[Peptide info]
ENST00000333725	549000331057	NP_996919.1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000343827	ENSP00000342459	NP_996923.1	[Transcript info]	[Exon info]	[Peptide info]

Features ▾

Chr 15
Length
Forward strand

Ensembl trans.
TCF12
Ensembl Known Protein Coding
NP_996919.1
Ensembl Known Protein Coding

DNA(contigs)
AC010999.6.1.221986
AC000002.1.1.198602

Orthologue Prediction

The following gene(s) have been identified as putative orthologues:

Species	Type	Gene identifier
<i>Clona savignyi</i>	1 to many	ENSCAVG00000011705 (Novel Ensembl prediction) [Info] [Align] No description
<i>Canis familiaris</i>	1 to 1	ENSCAFG000000016200 (TCF12) [Info] [Align] PROSITE: similar to Transcription factor 12 isoform 5 (Accession: NP_055491)
<i>Aedes aegypti</i>	1 to many	AAELI010226 (Novel Ensembl prediction) [Info] [Align] duph1er1es
<i>Bos taurus</i>	1 to 1	ENSBTAG00000002586 (TCF12) [Info] [Align] Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SwissProt, Acc: Q99061] [From human gene ENSG00000140262]
<i>Mus musculus</i>	1 to 1	ENSMUSG000000032228 (TCF12) [Info] [Align] transcription factor 12 (Source: MarkerSymbol, Acc: P901101877)
<i>Pan troglodytes</i>	1 to 1	ENSPTRG000000007109 (TCF12) [Info] [Align] No description

Ensembl v40: Homo sapiens Gene report for ENSG00000140262
Back Forward Reload Stop Home http://www.ensembl.org/Homo_sapiens/geneview?gene=ENSG00000140262;db=core;ENST00000343827 **Ensembl GeneView**

Transcript ENST00000343827

Transcript NP_996923.1 (RefSeq peptide ID). To view all Ensembl genes linked to the name [click here](#).

Transcript information
Exons: 13 Transcript length: 3,973 bps Protein length: 530 residues
[\[Further Transcript Info\]](#) [\[Exon Information\]](#) [\[Protein Information\]](#)

Similarity Matches
This Ensembl entry corresponds to the following database identifiers:
RefSeq peptide: NP_996923.1 [Target %id: 96; Query %id: 100] [\[align\]](#)
RefSeq DNA: NM_207040.1 [Target %id: 100; Query %id: 99] [\[align\]](#)
UniProtKB/TrEMBL: Q7Z3D9_HUMAN [Target %id: 100; Query %id: 100] [\[align\]](#)
Q9NQY2_HUMAN [Target %id: 7; Query %id: 100] [\[align\]](#)
Q9NQY4_HUMAN [Target %id: 4; Query %id: 100] [\[align\]](#)
Q9NQY5_HUMAN [Target %id: 4; Query %id: 100] [\[align\]](#)
Q9NQY6_HUMAN [Target %id: 9; Query %id: 100] [\[align\]](#)
Q9NQY7_HUMAN [Target %id: 6; Query %id: 100] [\[align\]](#)
6938
EntrezGene: A_14_P120408 [Target %id: 1; Query %id: 100]
Agilent CGH: A_23_P151930 [Target %id: 1; Query %id: 100]
A_24_P92142 [Target %id: 1; Query %id: 100]
EMBL: AF271611 [align] AF271612 [align] AF271613 [align] AF271614 [align]
AF271616 [align] AF271617 [align] BX537967 [align]
IP: IP00384979.2 [Target %id: 96; Query %id: 100]
Protein ID: AAF82573.1 [align] AAF82572.1 [align] AAF82573.1 [align] AAF82575.1 [align]
AAF82576.1 [align] AAF82578.1 [align] CAD97931.1 [align]
UniGene: Hs.595728 [Target %id: 9; Query %id: 99]
Hs.608843 [Target %id: 12; Query %id: 99]
Affymx Microarray Focus: Z08986_at
Affymx Microarray HuGeneFL: M83233_at
Affymx Microarray U133: Z08986_at 209986_at 209986_at 238041_at
Z38041_at Hs.21704.0_S2_3p_at Hs.82572.0_A1_3p_at
Affymx Microarray U95: 33348_at 33348_at 54311_at
Illumina: GL_4585865 [Target %id: 1; Query %id: 100]

GO
The following GO terms have been mapped to this entry via UniProt and/or RefSeq:
GO:0005634 [nucleus] IEA
GO:003028 [transcription regulator activity] IEA
GO:004549 [regulation of transcription] IEA

InterPro
IPRO01092 Basic helix-loop-helix dimerisation region bHLH - [\[View other genes with this domain\]](#)

Protein Family
ENSF00000000830 : TRANSCRIPTION FACTOR TRANSCRIPTION FACTOR
This cluster contains 3 Ensembl gene member(s) in this species.

Transcript structure

Protein features
Peptide

http://www.ensembl.org/Homo_sapiens/geneview?gene=ENSG00000140262;db=core;ENST00000343827

Ensembl v40: Homo sapiens Exon Report for ENST00000343827
Back Forward Reload Stop Home http://www.ensembl.org/Homo_sapiens/exonview?transcript=ENST00000343827;db=core **Ensembl ExonView**

Archival sites
Trace server
Sanger

Gene	Exon	Start	End	Strand	Length	Score
ENSE00001126370	15	55,342,602	55,342,764	1	2	163
Intron 10-11	15	55,342,765	55,352,519	-	-	9,755
ENSE00001126365	15	55,352,520	55,352,752	2	1	23
Intron 11-12	15	55,352,753	55,361,934	-	-	9,182
ENSE00001103528	15	55,361,935	55,362,088	1	-	154
Intron 12-13	15	55,362,089	55,365,646	-	-	3,558
ENSE00001405489	15	55,365,647	55,368,004	-	-	2,358

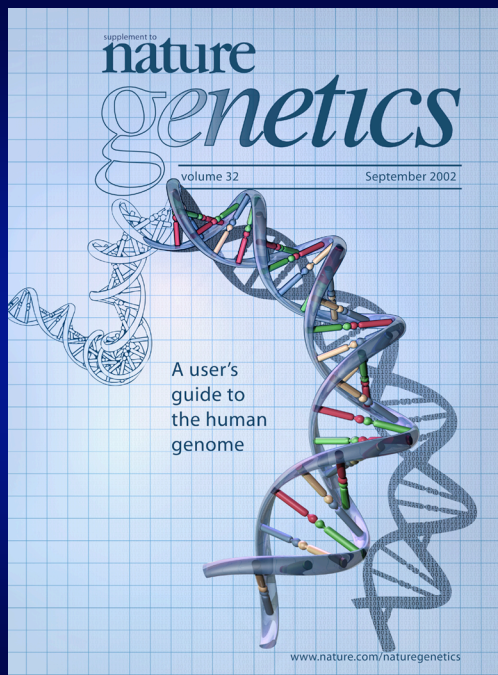
Supporting Evidence
The supporting evidence below consists of the sequence matches on which the exon predictions were based and are sorted by alignment score. There are a large number of supporting evidence hits for this transcript. Only the top ten hits have been shown. [Click to view all 16 supporting evidence hits.](#)

Score: 100 >=98 >=97 >=96 >=95 >=94 >=93 >=92 >=91 >=90 >=89 >=88 >=87 >=86 >=85 >=84 >=83 >=82 >=81 >=80 >=79 >=78 >=77 >=76 >=75 >=74 >=73 >=72 >=71 >=70 >=69 >=68 >=67 >=66 >=65 >=64 >=63 >=62 >=61 >=60 >=59 >=58 >=57 >=56 >=55 >=54 >=53 >=52 >=51 >=50 >=49 >=48 >=47 >=46 >=45 >=44 >=43 >=42 >=41 >=40 >=39 >=38 >=37 >=36 >=35 >=34 >=33 >=32 >=31 >=30 >=29 >=28 >=27 >=26 >=25 >=24 >=23 >=22 >=21 >=20 >=19 >=18 >=17 >=16 >=15 >=14 >=13 >=12 >=11 >=10 >=9 >=8 >=7 >=6 >=5 >=4 >=3 >=2 >=1 NO EVIDENCE

1. NM_207040.1
2. gi4037008|ref|NM_207040.1| Homo sapiens transcription factor 12 (TF12, helix-loop-helix transcription factor 12)
3. NM_207037.1
4. gi4037008|ref|NM_207037.1| Homo sapiens transcription factor 12 (TF12, helix-loop-helix transcription factor 12)
5. NM_207036.1

Additional resources

- UCSC Human Genome Browser User Guide
<http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html>
- NCBI Genomic Biology
<http://www.ncbi.nih.gov/Genomes/>
- NCBI MapViewer Help
<http://www.ncbi.nlm.nih.gov/mapview/static/MapViewerHelp.html>
- Ensembl Worked Example
http://www.ensembl.org/info/worked_example.pdf



<http://www.nature.com/ng/supplements/>

References

- Current Protocols in Bioinformatics
UNIT 1.4: The UCSC Genome Browser
UNIT 1.5: Using the NCBI Map Viewer to Browse Genomic Sequence Data
Access through <http://nihlibrary.nih.gov/ResearchTools/OnlineJournals.htm>
- UCSC
Hsu F *et al.* The UCSC Known Genes. *Bioinformatics*. 2006. 1034-46.
Hinrichs AS *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* 2006. 34:D590-8.
Kent WJ *et al.* Exploring relationships and mining data with the UCSC Gene Sorter. *Genome Res.* 2005. 15:737-41.
Hsu F *et al.* The UCSC Proteome Browser. *Nucleic Acids Res.* 2005. 33:D454-8.
Karolchik D *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004. 32:D493-6.
Karolchik D *et al.* The UCSC Genome Browser Database. *Nucleic Acids Res.* 2003. 31:51-4.
- NCBI
Wheeler DL *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2006:D173-80.
Dombrowski SM and Maglott D. Using the Map Viewer to Explore Genomes. *in* The NCBI Handbook. 2003.
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books>
- Ensembl
Birney E *et al.* Ensembl 2006. *Nucleic Acids Res.* 2006. 34:D555-61.
Hammond MP, and Birney E. Genome information resources - developments at Ensembl. 2004. *Trends Genet.* 20:268-72.
Birney E *et al.* An overview of Ensembl. 2004. *Genome Res.* 14:925-8.