


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/>

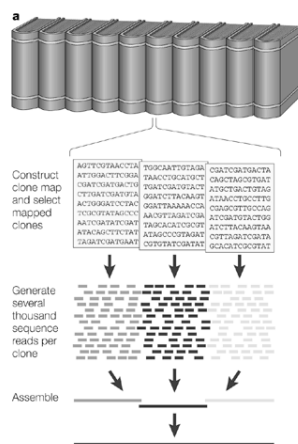
Ensembl
<http://www.ensembl.org>

Types of data integrated in genome browsers

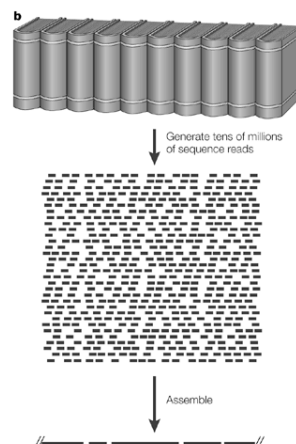
- Same starting material for all genome browsers: genomic sequence
- Annotations calculated independently by each genome browser
 - Genes
 - RefSeq mRNAs (non-redundant)
 - GenBank mRNAs (redundant)
 - ESTs
 - Gene predictions
 - SNPs
 - Homologous sequences from other organisms
 - STSs

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				

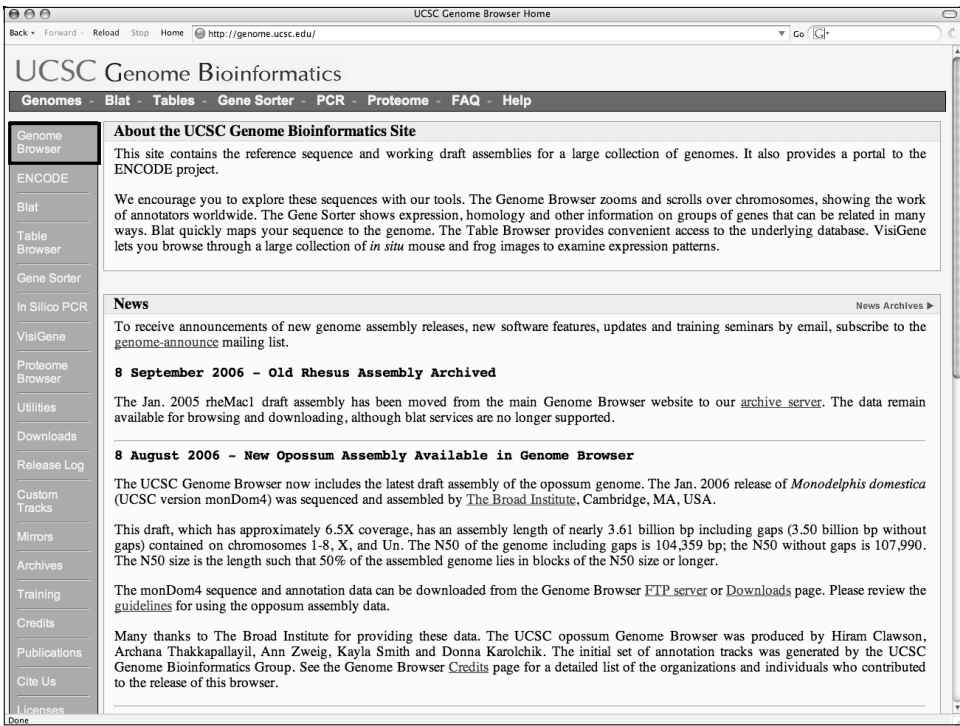
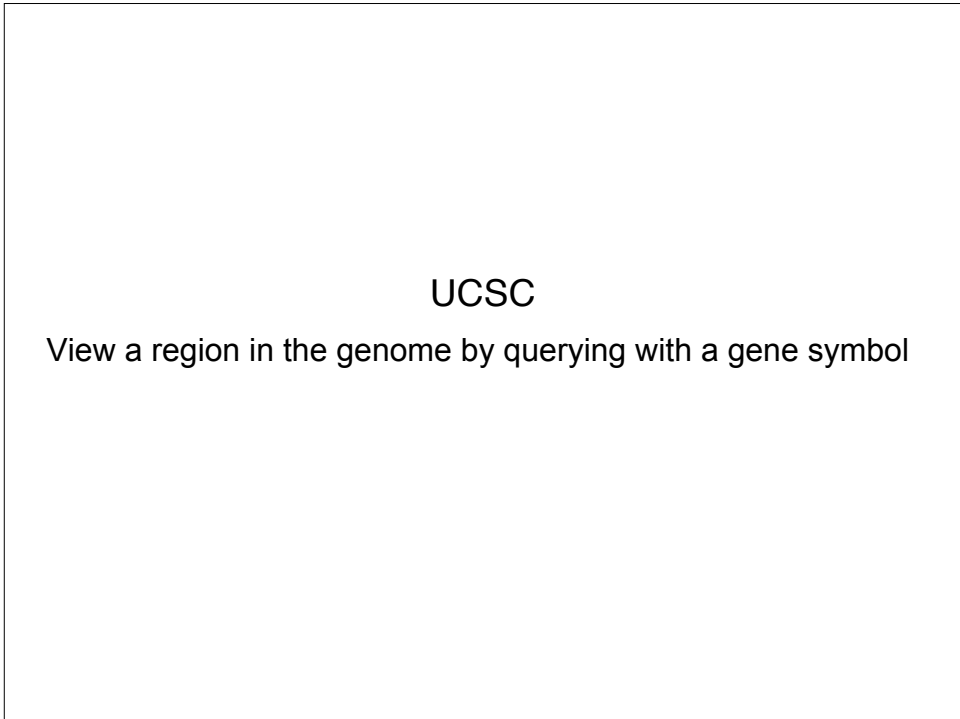
<http://www.ncbi.nlm.nih.gov/RefSeq/key.html>

```

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.
  1 (bases 1 to 1793)
REFERENCE   1 (bases 1 to 1793)
AUTHORS    Pappenberg, G., 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    14762366
.....
REFERENCE   134 (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    724701
COMMENT    REVIEWED REFSEQ: This record has been curated by NCBI staff. The
  reference sequence was derived from U00381.1 and M13132.1.
  On Jun 8, 1999 this sequence version replaced gi:4501885.

  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.
  COMPLETE REFSEQ complete on the 3' end.
.....
CDS         74..1791
            /gene="ACTB"
            /gc_component="actin filament; cytoskeleton; TIP60 histone
            acetyltransferase complex [pmid 10965108]"
            /gc_function="ATP binding; nucleotide binding; protein
            binding [pmid 15527821]; skeletal_muscle; cytoskeleton;
            cytoskeleton"
            /note="beta cytoskeletal actin; PS17P5-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:00032"
            /db_xref="MIM:102610"
            /translation="MDDIALLVVDSGCKAGFAGDADPAVFPISIVGRPHQVVM
            VMGQKDYVGDRAQSKRGLZLKYPIERGIIVNWDMEKIMHTFYNELRVAPEDIP
            VETLAPLAPFANRERKFTIMETFTWPAWVAQAVLELRASGTCIWNDSGQVY
            HTPVLYDGYALPHAILRLDLACRDLZYLKMLITRQVSPFTTAREIVRDKELKLY
            VALDFQSMATAASSSLSKVELDQGVTTIGNDRFPCPALPQPSFLMSSCQTHE
            FTNSLHKCVDIRKLYAVVLSGCTVYPCIALDQWKE;TALAFKRWKIIAPPD
            RKYSVWIGGSLASLSTFQQWISKQYVDSGSPSIVHKCF"
            ..
ORIGIN      1  ccgctcgcgc  ccgcagcacc  agacgctcgc  cttggccgat  ccgcgcgccy  tccacaccy
          61  ccgcccgcgc  accatcggag  atgatctacc  cgcgcctcgc  ctgcacagcg  gctccggcat
          121  gtcgaagcgc  ggtctcggcg  gcaacagatg  cccccggcgc  gttctccctc  ceatcgtggg
          181  ccgccaccag  caaccagggc  tgatcgtggg  catggctcac  aagatctcct  atcgtggcga
          241  ccaggcccaq  aqcaagagag  gcatctctac  cctcagatcc  cccatcagac  acggatcgt
          241
  
```

Beta actin mRNA RefSeq



Human (Homo sapiens) Genome Browser Gateway

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

Home Genomes Blat Tables Gene Sorter PCR FAQ Help

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: 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

About the Human Mar. 2006 (hg18) assembly

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

Sample position queries

A genomic 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.finger Lists many zinc.finger 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&hgslid=7752169

Known Genes

ADAM2 (NM_001464) at chr8:39720414-39814936 - ADAM metalloproteinase domain 2 proprotein
 ADAM2 (BC064547) at chr8:39720414-39814886 - ADAM2 protein.
 ADAM2 (BC034971) 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 chr14:69949574-69956774 - ADAM metalloproteinase domain 21 preproprotein
 ADAM23 (NM_003812) at chr2:207016613-207190922 - ADAM metalloproteinase domain 23 preproprotein
 ADAM20 (AF029899) at chr14:70058832-70061255 - ADAM metalloproteinase domain 20
 ADAM22 (BC034692) at chr7:87401638-87664383 - ADAM22 protein.
 ADAM22 (AF155381) at chr7:87401638-87664383 - 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 chr14:69949574-69956774 - (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
 ADAM23 at chr2:207016613-207190922 - (NM_003812) ADAM metalloproteinase domain 23 preproprotein
 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 chr8:39723146-39814902 - (NM_213957) fertilin beta
 ADAM2 at chr8:39723147-39813888 - (NM_17428) a disintegrin and metalloproteinase domain 2
 ADAM2 at chr8:39723147-3981569 - (NM_009618) a disintegrin and metalloproteinase domain 2
 ADAM2 at chr8:39723147-39813877 - (NM_020077) a disintegrin and metalloproteinase domain 2
 ADAM21 at chr14:69949574-69956774 - (NM_020330) a disintegrin and metalloproteinase domain 21
 ADAM21 at chr14:69949574-70061197 - (NM_020330) a disintegrin and metalloproteinase domain 21
 ADAM22 at chr7:87401747-87649282 - (NM_001007220) a disintegrin and metalloproteinase domain 22
 ADAM22 at chr7:87401747-87649282 - (NM_001007221) a disintegrin and metalloproteinase domain 22
 ADAM23 at chr2:207016603-207194450 - (NM_011780) a disintegrin and metalloproteinase domain 23
 ADAM23 predicted at chr2:207016759-207190724 - (NM_001029899) a disintegrin and metalloproteinase domain 23
 ADAM24 at chr8:17371373-17373360 - (NM_010086) a disintegrin and metalloproteinase domain 24
 ADAM24 at chr8:188904470-188906353 - (NM_010086) a disintegrin and metalloproteinase domain 24
 ADAM24 at chr14:70059197-70060969 - (NM_010086) a disintegrin and metalloproteinase domain 24
 ADAM25 at chr14:188904411-188906248 - (NM_011781) a disintegrin and metalloproteinase domain 25
 ADAM25 at chr14:188904470-188906353 - (NM_011781) a disintegrin and metalloproteinase domain 25
 ADAM25 at chr14:69949072-69955994 - (NM_011781) a disintegrin and metalloproteinase domain 25
 ADAM25 at chr14:69949072-69955994 - (NM_011781) a disintegrin and metalloproteinase domain 25
 ADAM25 at chr14:69949072-69955994 - (NM_011781) a disintegrin and metalloproteinase domain 25
 ADAM26 at chr8:17371373-17373360 - (NM_010085) a disintegrin and metalloproteinase domain 26
 ADAM26 at chr8:17371373-17373360 - (NM_010085) a disintegrin and metalloproteinase domain 26

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

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	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	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					

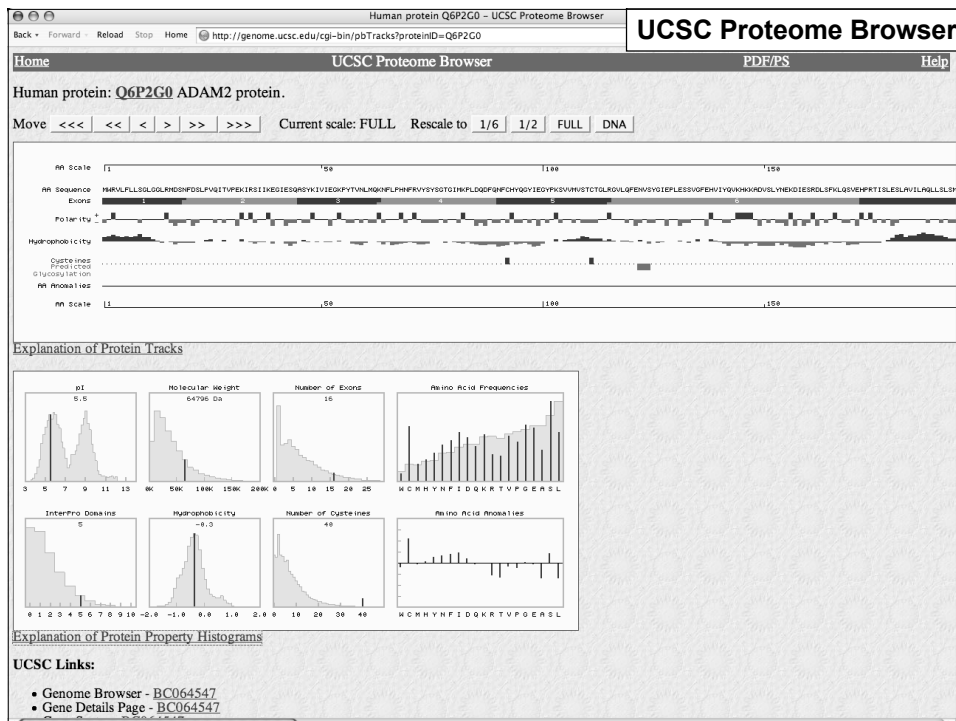
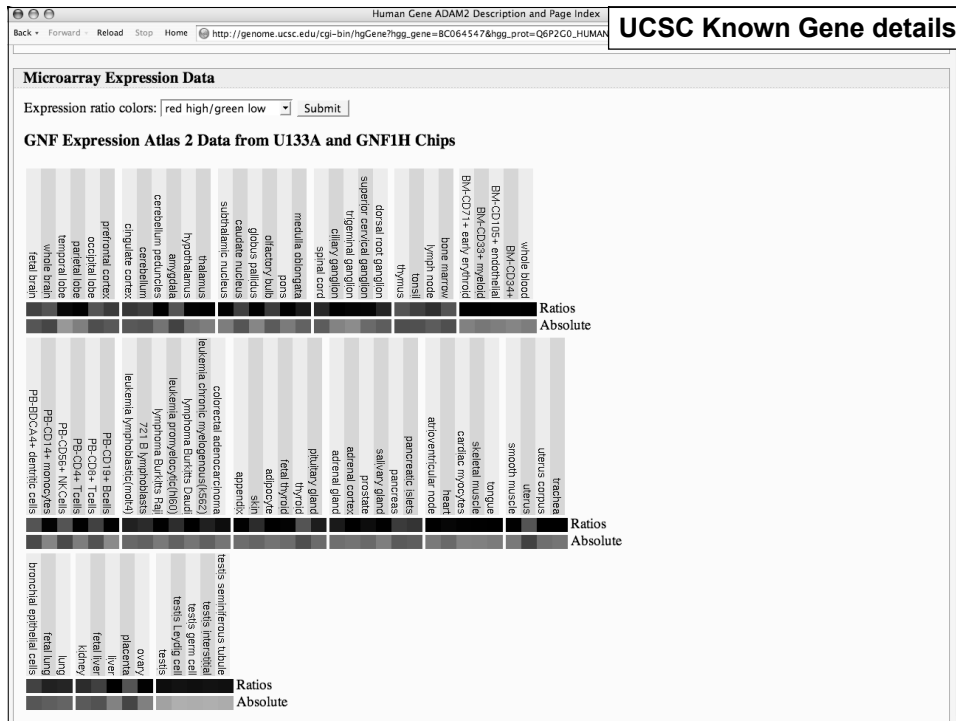
Sequence

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



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

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

Home Genomes Blast 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

RefSeq Gene

UCSC RefSeq Gene details

RefSeq Gene ADAM2

RefSeq: [NM_001464.3](#) Status: **Reviewed**
 CDS: 3 complete
 OMIM: [601533](#)
 Entrez Gene: [2515](#)
 PubMed on Gene: [ADAM2](#)
 PubMed on Product: [ADAM metallopeptidase 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](#)
- [mRNA Sequence](#) *different from the genomic sequence.*
- [Genomic Sequence](#) *from assembly*

click

Genomic Sequence Near Gene

UCSC RefSeq Gene details

Home Genomes Genome Browser Blat Tables Gene Sorter PCR FAQ Help

Genomic Sequence Near Gene

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
atatacacagtagaatacctactattacacccaaagcattgagaaga
atgtttatagctaaatatttttaaatagctggaacataaacaacaa
aatattcatcaacgtaaaatggaacacaaagtggttatattatga
atittgaatatacaccaatgaggataaacagaactatgtttgatga
accttacaatcatctctattaaagaacccagactgaaaggtatggtt
gattgcttctacttgcgaaaagtcaaaaacagacaaaacgaatcttgg
ttgtttagaagtcactggttgaggttggaaatcggggatttgggtgg
ctttttcatctctcaactggtgactagttaactggtttttttttccac
ttgaatattaatgaactgtgaacttatgatttatataacttttttc
gtttttgttctctcttttttttttttttttttttttttttttttttt
tttgtctctcaaccaggctggagtgcaagtgtaaggtctctctgctca
tgcacctctgctctaggttcaagcattctctgctcagctctcccg
agttagctgggatttcaaggacccgcaacatgctggtaatttttttt
gtatttttagtacaaggggtttcaacatttgggaaactggtctoga
actcctgacgtgtttatataattcaattgaaatttacttaagaagt
gtttataaattctctgttctcagcttggaaagtattttgttgg
tgttgcttaattaggatcaactcagtggaagtgtctgtctcgaagag
acaggtctcaggaactcgaacttccacagcaccacacacacactcag
cccactgggtctcccagcgcctacctcttccagcgtgctggccggg
```

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 (311,220) [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 HuEx 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 pack	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=93E3E+&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 (311,220) [39,765,312 39,859,834]

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.

Mapping and Sequencing Tracks

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

The screenshot shows a web browser window with the URL <http://genome.ucsc.edu/cgi-bin/hgCtHgsid=77541763&o=39844165&t=39844181&g=tfbsC>. The page title is "UCSC TFBS Track details". The browser's address bar shows the URL. The page content includes a navigation menu with links for Home, Genomes, Genome Browser, Blat, Tables, Gene Sorter, PCR, FAQ, and Help. The main heading is "HMR Conserved Transcription Factor Binding Sites (V\$HMEF2_Q6)". Below this is a section titled "Transcription Factor Binding Site information:" with the following details: Item: V\$HMEF2_Q6, Transfac matrix link: M00406, Score: 874, zScore: 3.50, Strand: -, Position: chr8:39844166-39844181, Band: 8p11.22, Genomic Size: 16, and a link to "View DNA for this feature". A second section is titled "Transcription Factors known to bind to this site:" and lists Factor: MEF-2A, Species: human, SwissProt ID: Q02078, and a link to "Proteome Browser Entry". It also includes a link to "View table schema" and "Data last updated: 2004-12-14". A "Description" section explains that the track contains the location and score of transcription factor binding sites conserved in the human/mouse/rat alignment. It notes that a binding site is considered conserved 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. It also states that in the graphical display, each box represents one conserved tfbs, and 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. Finally, it mentions that 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

UCSC Genome Browser on Human May 2004 Assembly

position/search chr8:39,765,312-39,859,834

Click

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	WSSD Duplication	Short Match	Restr Enzymes

Extended DNA Case/Color Options

Use this page to highlight features in genomic DNA text. DNA covered by a particular track below for details about color, and for examples. Tracks in 'hide' display mode are not shown.

Position chr8:39,765,312-39,859,834 Reverse complement

Letters per line 60 Default case: Upper Lower submit

Track Name	Toggle Case	Underline	Bold	Italic	Red	Green	Blue
Chromosome Band (Ideogram)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
STS Markers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
Known Genes	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	255	0	0
RefSeq Genes	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	255	0	0
MGC Genes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
Vega Genes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
Vega Pseudogenes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
Exoniphy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
Human mRNAs	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	255	0	0
Spliced ESTs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
ExonWalk	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
TFBS Conserved	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	255	0
Conservation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
SNPs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
RepeatMasker	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0

```

tgaagtggaaaaccctcaggaaagcctctccaggatgtcaatgtaacttggaggcaa
aggggagtagcgtgaggtcccaaaagcagagaggggtttttctgtacttctg
TCCATCCGACGCCCGCCGACCCGCTGAGCAGAAACAAGCCGCCACATGCTTGAAGT
CCTGGGTCCACCCGGAATAAT GGTTACA GAGCGCGAAT
GACGCCCCgcaacgcgctggaagaggtaggcgtgagcgcctggagcctggaggtgag

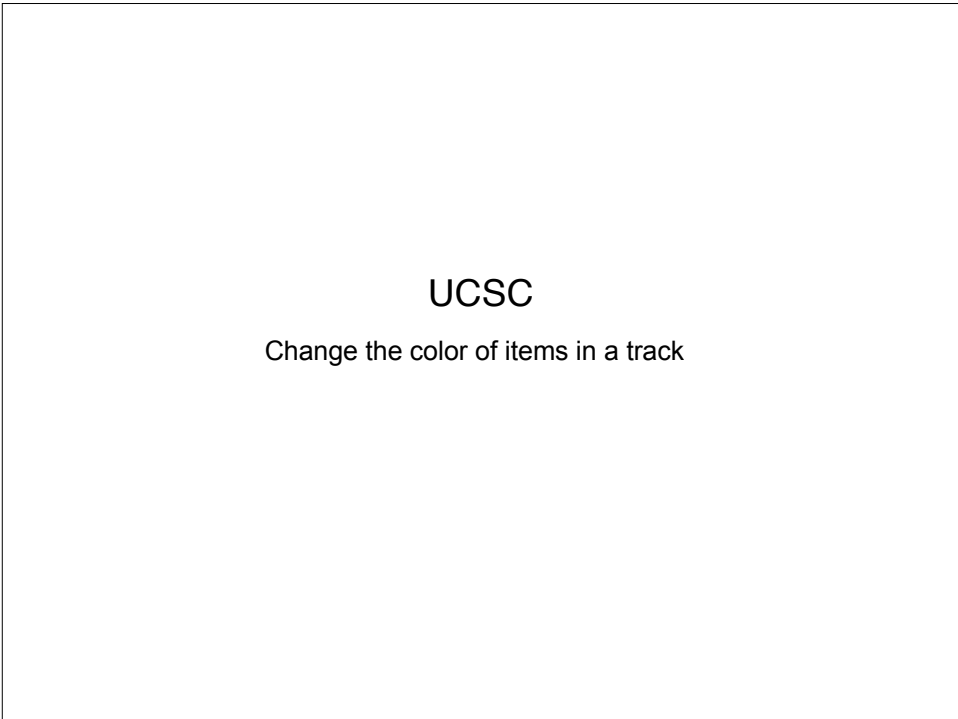
```

```

cttttcaactcaactctgaagtgtttctctctctctctctctctctctctctctct
ctgtctctctctctctctctctctctctctctctctctctctctctctctctctct
ttgttcaagcatttgtgtatgacattctctctctctctctctctctctctctctct

```

Red: mRNA sequences
Green: Transfac TFBS
Yellow: mRNA + TFBS



The screenshot displays the UCSC Genome Browser interface for Human chromosome 8, region 39,720,414-39,814,936. The browser title is "UCSC Genome Browser on Human Mar. 2006 Assembly". The URL is "http://genome.ucsc.edu/cgi-bin/hgTracks".

Navigation controls include "move <<< << < > >> >>>" and "zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x". A search bar contains "chr8:39,720,414-39,814,936" with "jump clear size 94,523 bp. configure" options.

The main visualization area shows several tracks:

- chr8: 39750000 39850000
- UCSC known Genes Based on UniProt, RefSeq, and GenBank entries
- RefSeq Genes
- Human ESTs from GenBank
- Human ESTs That Have Been Spliced
- Spliced ESTs
- Conservation (17 Species)
- RepeatMasker
- Simple Repeats (copy# 1-100)

Below the tracks are navigation buttons: "move start < 2.0 >" and "move end < 2.0 >". A central instruction reads: "Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options."

Additional controls include "default tracks hide all add custom tracks configure refresh". A note states: "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."

The "Mapping and Sequencing Tracks" section includes:

Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb Rate
dense	hide	hide	hide	hide

The "Variation and Repeats" section includes:

SNPs	Segmental Dups	RepeatMasker	Simple Repeats	Microsatellite
dense	hide	dense	hide	hide

Other controls include "Self Chain hide" and a "refresh" button.

A black arrow labeled "click" points to the "SNPs" track in the "Variation and Repeats" section.

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:
 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

UCSC SNP Track

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

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

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

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

UCSC

Find a chicken homolog of a human protein

The screenshot displays the NCBI Entrez Protein search results for the query 'np_001455'. The search results show one entry: '2 proprotein [Homo sapiens] (5743080)'. The protein sequence is displayed in FASTA format, starting with 'MWRVLLSGLGCLMDSNFDLSPVQIVPKIRASIKRQIESQASYKIVIRGKPYVNLQKMFPHNP...'. The search interface includes a search bar, navigation buttons (Back, Forward, Reload, Stop, Home), and a sidebar with various tools and links.

```
>I: NP_001455. Reports ADAM metallopepti...[gi:55743080]
>gi|55743380|ref|np_001455.3|ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
MWRVLLSGLGCLMDSNFDLSPVQIVPKIRASIKRQIESQASYKIVIRGKPYVNLQKMFPHNP
RVYSYSGTCMKPLDQDFQPCYQYISGYPKSVWVSSCTCLRGVLPENVSYGIDPLSSVGFEBVT
YQKRRKADVLEWSDISBLSFPLQGVKQGFYAVIDMVIWKGQLRNGSGDTPVAKQKPFELG
LWNAIFVSNITIISSLELWIDENIAPTGRANELLFTLWNTSVLVLRPHDVAFLVREKSNVGA
TPOGKMDANFAGGVLEPFTIISLESLAVLQGLSLSMGIITDIDINWQCSGAVCMNPERITFSQVKE
FENCSEDFPAREFISKQKQCLINQPLRDPFFKQAVCCNAKLEAGBCCDCTQDQCALIGETCCDIATCR
FRAGSCADQPCENCLPMSKRNCRPSFEBDCDLPETCNGSSASCPEMNIYVOTGHPCCGLNQWICIDQVCR
SGMKCTPTTQKVEVETGSECVHILKSKVSRKQCIISDGYVQCEADNLQCKLCKYVGFLLQIPRA
TIIANISGLCIAVDFASDHADQKHWIKDQTSQCSNVCNRQCVSSSYLVGDCCTDKCNDRQVCNKK
KMICGSIYLPFCVQGLDMPGSIIDGNFPFVAIPARLERRITENIYHKKRHWPFLLPFFIIFC
VLIAMVKNVNFQRKWRTEYSDSDQPSSESEPKG
```


Chicken BLAT Search

UCSC BLAT search

Home Genomes Tables Gene Sorter FAQ Help

Chicken BLAT Search

BLAT Search Genome

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

```
>gl|55743080|zef|NP_001455.3| ADAM metalloproteinase domain 2 proprotein [Homo sapiens]
MNRVFLLEGLGGLRMDGDFSLPVOITVPEKIRSLIKRIGESQASYKIVIRGKPYTVNLGKMFLEHNP
RVVSYSGTGIMKPLDDQFNFCYQVIGYPKSVVMVSTCTGLRGVLPQVNSVGIPLPSSVGFHEVI
YQVHKKADVSLYNEKDIERSDLSFKLQSVPEQQDFAKYIEMHIVKQLYNHMGSDTTVAQRVFLIG
LINAIPVSPNITLLSLELWIDENKIATTOGANELLFTPLRWTSYLVRPHDVAFLLVYREKSNVGA
FTQGRKWDVNYAGGVVLRHTISLRELAIVLQGLLELDMGTYVDLNMKCCGSAVCIHWPEAHPSSQVYI
FNSCSFDFAHFISKQKSQCLHNQPRLDPPFQQAQVGNKLEAGEECDCOTEDCALIGTCCDIATCR
FKAGSNCADGPCENCLFMSKERMCRPSFEBDCDLPFYCNGSSASCPENHYVQTHGPCGLNQCIDGVCN
SGDQCTPTFGKEVEFPDPSQVSYLMSKFTVSGNCCIISDGYTQCBADNLQCKLQKTYVGRFLQIPRA
TIIYANISGHLCAVEFASDHADSQRMWIKDQTSKGNKVCNRQCVSSVLYGDCCTDKCNDRGVCNNK
KHCCHCSAYLPDQSVQSLDWPGGSIDSGNFPVVAIPARLPERRYENIYHSKPRWPFLLPFPFFIIFC
VLIAMVKNVFORRWRTEDYSSDQPESESEPKG
```

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.

Chicken BLAT Results

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	6765972	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

Chicken BLAT Results
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	++	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 in sequence.

NP_001455.3

```

mwrwlflag leglrmdanf dslprvqitvp ekirsiikeg leeqaaykiv leqkpytvl 60
mqknflphnf rvyssyagti mkpldqdfn fchygyyieg ypkavvmvat etqirgvlqf 120
enveyyiepl eavvygiehvi yqvkhhkadv elynekdies rdlsfkigev epqgdfakyl 180
emvsvveqel yhmegsdty veqvigilg llnaifvein lillisaapl widentiaat 240
qeanelhbf lrwktaylwl rphdvaillv yreksnyva tfgqmodan yaggvvlhpr 300
tislalevli laqlisimg ltyddinkqg egeavcimp eaahfsgvki fanesiedia 360
hfiakqkqgc lhnqprldpf lkqavogna klageeecd gteqdcailg etocdiator 420
fkagncsaeg pccencifms kormerpefe eodlpeycng saascpenhy vtqghpegin 480
gdcidavem sgdqctdtdf gkeveropse eyahinaktd vagnocisda vylceaad 540
qCKLlCk v qkllqipra IIVAnisgl Llavefaad hadsqkwi DGTaCcaNk 600
crngrevasa ylgpdeitdk endgvcmsk khheasayl pddcveqadl wpggsidagn 660
fppvalparl perzyienly hskpmrvpff lfipffiiic vliainvkn lqkkrwtd 720
yasdeqpeae sepkq
                    
```

Chicken.chrUn :

```

AACCTGGGCT GTGGAAACT CACTGCaCa CA'ccaaac ggttccctt caccacaata 635429
aagggt CCA TCATCATGC TcaagtCaCa gaacACTCTT C ggttcttt lqkgtlaag 635489
catgaacct ceggsacaga tctctctctg gttAGGATG GCACGaaATC CGGTceegga 635549
AAGCT:
                    
```

Side by Side Alignment*

```

001615 N L O C G K L I C K Y 001647
>>>>> | | G | | | | | | T | >>>>>
635370 aatctgggtgtggaaaactcatctgcacatac 635402

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

001769 K D G T S C S N K V 001809
>>>>> | | | | | K | | P G | >>>>>
635523 aaggatggcagaatgggtcccggaagta 635555
                    
```

UCSC

Add your own custom tracks

18

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=chrXG3A15192C073K2C054-151X Co

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 38966888 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 intersect. 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 out)
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](#)

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	#filter: refGene.exonCount = 1															
2	#bin	name	chrom	strand	txStart	txEnd	cdsStart	cdsEnd	exonCount	exonStarts	exonEnds	id	name2	cdsStartStat	cdsEndStat	exonFrames
3	88	NM_009511	chr1	+	15858950	15860803	15858950	15860803	1	15858950	15860803	0	B5C1A1	cmpl	incmpl	0
4	178	NM_002232	chr1	-	11015832	11019378	11017226	11018954	1	11015832	11019378	0	KCNM3	cmpl	cmpl	0
5	301	NM_001321	chr1	-	23983790	23986585	23986520	23986581	1	23983790	23986585	0	CHML	cmpl	cmpl	0
6	585	NM_001005484	chr1	+	58953	59871	58953	59871	1	58953	59871	0	OR4F5	cmpl	cmpl	0
7	587	NM_001005221	chr1	+	357521	358458	357521	358458	1	357521	358458	0	OR4F29	cmpl	incmpl	0
8	587	NM_001005224	chr1	+	357521	358458	357521	358458	1	357521	358458	0	OR4F3	cmpl	incmpl	0
9	587	NM_001005277	chr1	+	357521	358458	357521	358458	1	357521	358458	0	OR4F16	cmpl	incmpl	0
10	589	NM_001005221	chr1	-	610960	611897	610960	611897	1	610960	611897	0	OR4F29	incmpl	cmpl	0
11	589	NM_001005224	chr1	-	610960	611897	610960	611897	1	610960	611897	0	OR4F3	incmpl	cmpl	0
12	589	NM_001005277	chr1	-	610960	611897	610960	611897	1	610960	611897	0	OR4F16	incmpl	cmpl	0
13	593	NM_008605	chr1	+	1157507	1160281	1157521	1158311	1	1157507	1160281	0	BDGAT6	cmpl	cmpl	0
14	607	NM_080431	chr1	+	2927905	2929325	2928110	2929244	1	2927905	2929325	0	ACTR12	cmpl	cmpl	0

NCBI

View a genomic region between two STS markers

NHGRI Current Topics in Genome Analysis 2006

Mining Genomic Sequence Data

NCBI Home Page
National Center for Biotechnology Information
National Library of Medicine, National Institutes of Health

Search | All Databases | BLAST | OMIM | Books | TaxBrowser | Structure

Search [] for [] Go []

SITE MAP
Alphabetical List
Resource Guide

About NCBI
An introduction to NCBI

GenBank
Sequence submission support and software

Literature databases
PubMed, OMIM, Books, and PubMed Central

Molecular databases
Sequences, structures, and taxonomy

Genomic biology
The human genome, whole genomes, and related resources

Tools
Data mining

Research at NCBI
People, projects, and seminars

Software engineering
Tools, R&D, and databases

Education
Teaching resources and on-line tutorials

FTP site

What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)

Hot Spots
 ▶ 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
 ▶ dBMHC
 ▶ Mouse genome resources
 ▶ My NCBI
 ▶ ORF finder
 ▶ Rat genome resources
 ▶ Reference sequence project
 ▶ SAGEmap
 ▶ SKY/CGH database
 ▶ dbSNP

Whole Genome Association
The NCBI Whole Genome Association (WGA) resource provides researchers with access to genotype and associated phenotype information that will help elucidate the link between genes and disease. For more information, click here to see the [WGA](#) resource page and click here to read the [press release](#).

100 Gigabases
GenBank and its collaborating databases, the European Molecular Biology Laboratory and the DNA Data Bank of Japan, have reached a milestone of 100 billion bases from over 165,000 organisms. See the [press release](#) or find more information on [GenBank](#).

PubMed Central
An archive of life sciences journals
 • Free fulltext
 • Over 500,000 articles from over 200 journals
 • Links to PubMed and fully searchable
 Use of PubMed Central requires no registration or fee. Access it from any computer with an internet connection.

NCBI News
Summer 2006 News available online

Map Viewer
http://www.ncbi.nlm.nih.gov/mapview/

Genomic Biology | Genome | Taxonomy | Entrez | BLAST | Help

Search [Homo sapiens (human) Build 36] for [d8s1170 OR d8s94] Go []

Click on the organism name to go to the genome view

Vertebrates

Mammals
 BLAST *Bos taurus* (cow)
 BLAST *Canis familiaris* (dog)
 BLAST *Felis catus* (cat)
 BLAST *Homo sapiens* (human) Build 36
 BLAST *Homo sapiens* (human) Build 35
 BLAST *Macaca mulatta* (rhesus macaque)
 BLAST *Mus musculus* (mouse) Build 36
 BLAST *Mus musculus* (mouse) Build 35
 BLAST *Ovis aries* (sheep)
 BLAST *Pan troglodytes* (chimpanzee)
 BLAST *Rattus norvegicus* (rat)
 BLAST *Sus scrofa* (pig)

Other Vertebrates
 BLAST *Danio rerio* (zebrafish)
 BLAST *Gallus gallus* (chicken)

Invertebrates

Insects
 BLAST *Anopheles gambiae* (mosquito)
 BLAST *Apis mellifera* (honey bee) Amel_4.0
 BLAST *Apis mellifera* (honey bee) Amel_2.0
 BLAST *Drosophila melanogaster* (fruit fly)
 BLAST *Tribolium castaneum* (red flour beetle)

Nematode
 BLAST *Caenorhabditis elegans* (nematode)

Echinoderms
 BLAST *Strongylocentrotus purpuratus* (purple sea urchin)

Protozoa
 BLAST *Cryptosporidium parvum*
 BLAST *Dictyostellium discoideum*
 BLAST *Plasmodium falciparum*

Plants BLAST Search all plant maps
 BLAST *Arabidopsis thaliana* (thale cress)
 BLAST *Avena sativa* (oat)
 BLAST *Beta vulgaris* (beet)
 BLAST *Glycine max* (soybean)
 BLAST *Hordeum vulgare* (barley)
 BLAST *Lotus japonicus* (lotus)
 BLAST *Lycopersicon esculentum* (tomato)
 BLAST *Manihot esculenta* (cassava)
 BLAST *Oryza sativa* (rice)
 BLAST *Triticum aestivum* (wheat)
 BLAST *Zea mays* (corn)

Fungi BLAST Search all fungal maps
 BLAST *Aspergillus fumigatus*
 BLAST *Candida glabrata*
 BLAST *Cryptococcus neoformans*
 BLAST *Debaryomyces hansenii*
 BLAST *Encephalitozoon cuniculi*
 BLAST *Eremothecium gossypii*
 BLAST *Gibberella zeae*
 BLAST *Kluyveromyces fragilis*
 BLAST *Magnaporthe oryzae*
 BLAST *Neurospora crassa*
 BLAST *Saccharomyces cerevisiae* (baker's yeast)
 BLAST *Schizosaccharomyces pombe* (fission yeast)
 BLAST *Ustilago maydis*
 BLAST *Yarrowia lipolytica*

See more about [Bacteria](#) (BLAST), [Organelles](#) (BLAST), [Viruses](#) (BLAST)

The Map Viewer supports search and display of genomic information by chromosomal position. Regions of interest can be retrieved by text queries (e.g. gene or marker name) or by sequence alignment (BLAST). View results at the whole genome level, and select what to display in more detail. Multiple options exist to configure your display, download data, navigate to related data, and analyze supporting information using the tools provided. [More...](#)

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&OR=d8s94

NCBI NCBI Map Viewer

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

PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help

Show related entries Help FTP Map Viewer home

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

BLAST search the human genome

1 2 3 4 5 6 7 8 9 10 11 12 13
14 15 16 17 18 19 20 21 22 23 Y MT

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

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

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

NCBI NCBI Map Viewer

Search BLAST OMM Taxonomy Structure

Human genome overview sw page (Build 36.2)
Human genome overview sw page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
[Maps & Options]
Compress Map
Region Shown:
32M
44,400K Go
out
zoom
in
You are here:
Ideogram
9p23
9p21
9p11
9p11.2
8q12

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: STS
Region Displayed: **32M-44,400K bp**
Map 1: UniGene Clusters
Map 2: Genes On Sequence
Map 3: STS

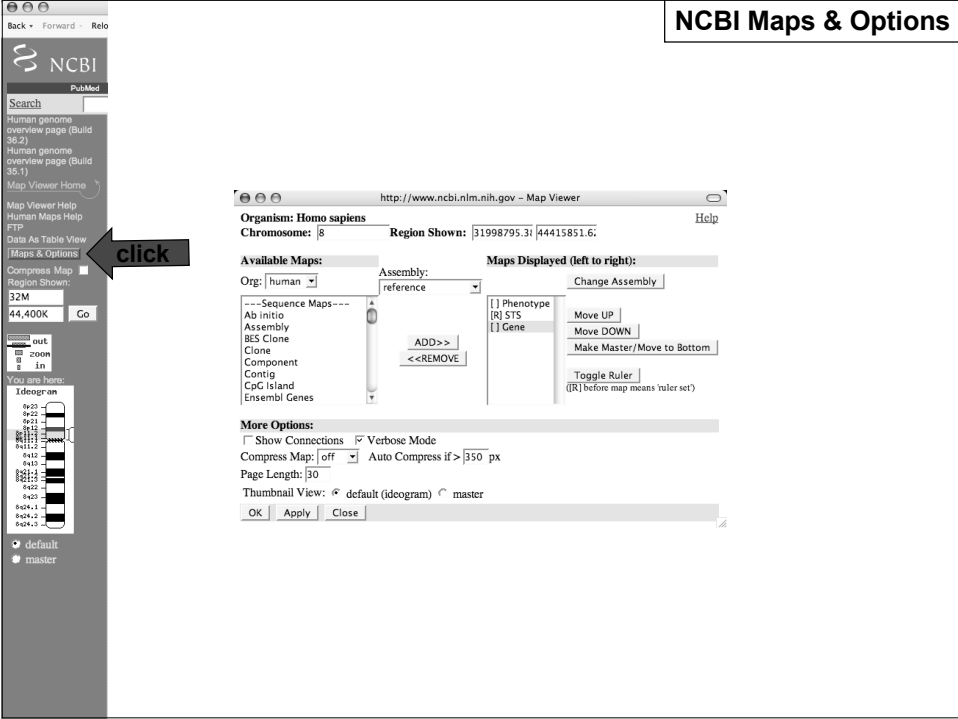
Marker	Go to
RH120054	
STS-N33323	
G29276	
D8S1170	
RH36053	
SHGC-110285	
SHGC-156087	
RH122187	
AFMB283XE1	
RCP_3679	
RH35988	
WL17448	
SHGC-79887	
CHNB3 2327	
G17210	
D8S94	
RH67241	

Summary of Maps:
Map 1: Homo sapiens UniGene Clusters Table View
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 Table View
Region Displayed: **32M-44,400K bp**
Total Genes On Chromosome: **984** (not localized)
Genes Labeled: **50** Total Genes in Region: **85**
Map 3: STS Table View
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

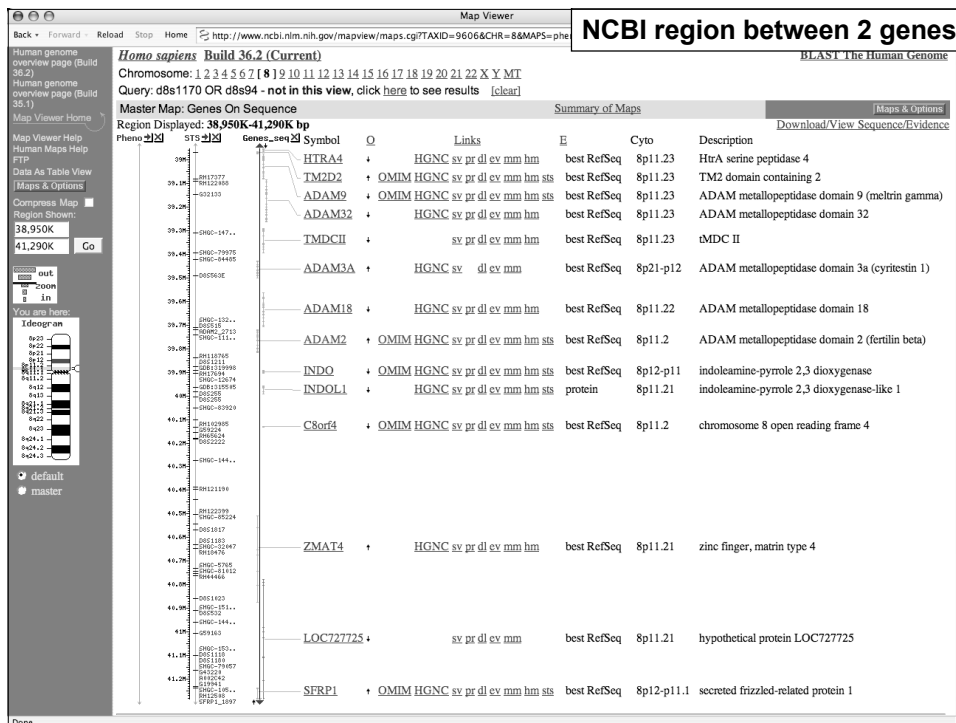
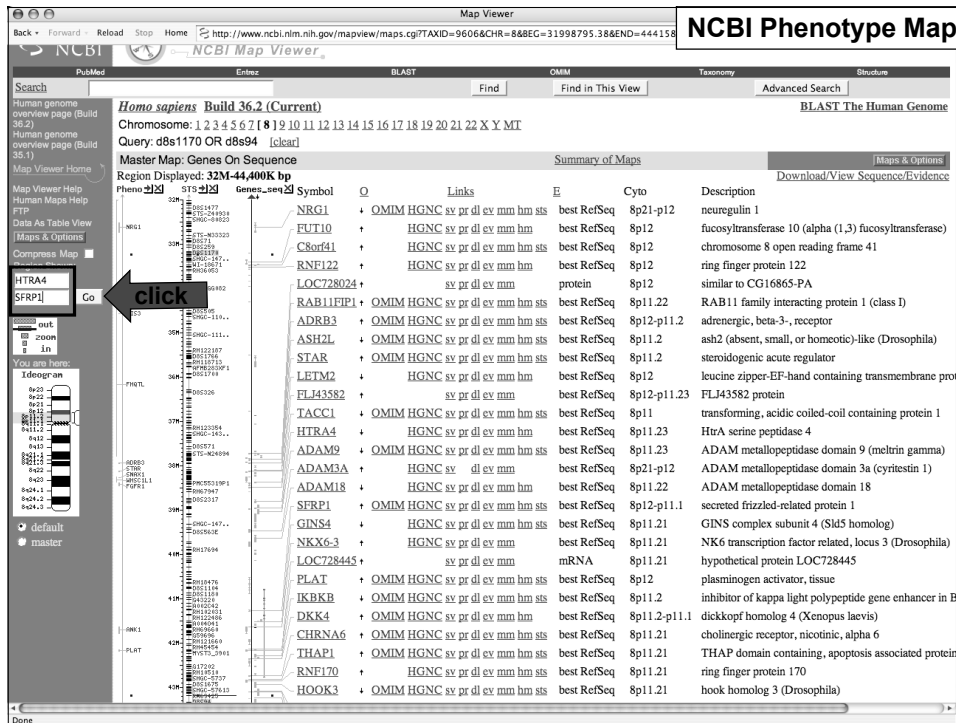
NCBI Maps & Options

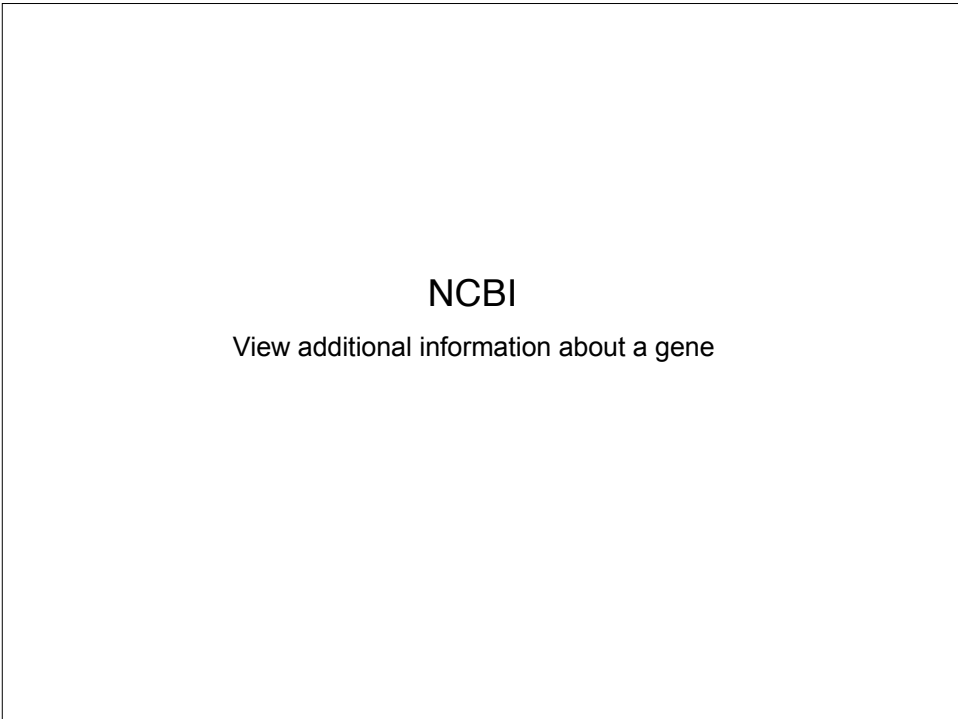


The screenshot displays the NCBI Map Viewer interface. On the left is a navigation sidebar with a search bar and various links. A black arrow labeled "click" points to the "Maps & Options" link in the sidebar. The main content area shows the "Map Viewer" window for Homo sapiens, with the region 31998795.31-44415851.6 selected. It features two columns: "Available Maps" and "Maps Displayed (left to right)". The "Available Maps" list includes items like "Ab initio", "Assembly", "BES Clone", "Clone", "Component", "Contig", "CpG Island", and "Ensembl Genes". The "Maps Displayed" list includes "Phenotype", "STS", and "Gene". Below these are "More Options" such as "Show Connections", "Verbose Mode", "Compress Map", "Page Length", and "Thumbnail View".

NHGRI Current Topics in Genome Analysis 2006

Mining Genomic Sequence Data





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 Download/View Sequence/Evidence

Region Displayed: 38,950K-41,290K bp

Pheno	STS	Gene	Symbol	Q	Links	E	Cyto	Description
39,148	BM12377	HTRA4	HTRA4	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.23	HtrA serine peptidase 4
39,148	BM12180	TM2D2	TM2D2	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.23	TM2 domain containing 2
39,206	BM12180	ADAM9	ADAM9	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.23	ADAM metalloproteinase domain 9 (meltrin gamma)
39,206	BM12180	ADAM32	ADAM32	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.23	ADAM metalloproteinase domain 32
39,206	BM12180	TMD2CII	TMD2CII	+	sv pr dl ev mm hm	best RefSeq	8p11.23	TMD2C II
39,442	BM12180	ADAM3A	ADAM3A	+	HGNC sv dl ev mm	best RefSeq	8p21-p12	ADAM metalloproteinase domain 3a (cytostin 1)
39,442	BM12180	ADAM18	ADAM18	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.22	ADAM metalloproteinase domain 18
39,776	BM12180	ADAM2	ADAM2	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.2	ADAM metalloproteinase domain 2 (fertilin beta)
39,776	BM12180	INDO	INDO	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p12-p11	indoleamine-pyrrole 2,3 dioxygenase
40,000	BM12180	INDOL1	INDOL1	+	HGNC sv pr dl ev mm hm sts	protein	8p11.21	indoleamine-pyrrole 2,3 dioxygenase-like 1
40,000	BM12180	C8orf4	C8orf4	+	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	8p11.2	chromosome 8 open reading frame 4
40,000	BM12180	ZMAT4	ZMAT4	+	HGNC sv pr dl ev mm hm	best RefSeq	8p11.21	zinc finger, matrin type 4
40,000	BM12180	LOC727725	LOC727725	+	sv pr dl ev mm	best RefSeq	8p11.21	hypothetical protein LOC727725
40,000	BM12180	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
All Databases: PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, Books, OMM

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
chromosome: 8; Location: 8p11.2
See ADAM2 in MapViewer

Bibliography Gene References into Function (GeneRIF): Submit

Interactions

Description

ADAM2 Product	Interactant	Other Gene	Complex	Source	Pubs
Done					

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-111629(c-PCR) (Links: [UniSTS:168466](#))
ADAM2_2713(c-PCR) (Links: [UniSTS:462012](#))

Gene Ontology
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	AAB34957
mRNA BC064547	AAB64547
mRNA BG719616	None
mRNA B1460477	None

OMIM - A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 2; ADAM2

Back Forward Reload Stop Home <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=601533> Co Clear

NCBI OMIM Online Mendelian Inheritance in Man Johns Hopkins University

Search OMIM for Co Clear

Limits Preview/Index History Clipboard Details

Display Detailed Show 20 Send to

All: 1 OMIM dBSNP: 0 OMIM UniSTS: 0

***601533** [Links](#)

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

Back Forward Reload Stop Home http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=homologene&dopt=HomoloGene&list_uids=1127 My NCBI [Sign In] [Register]

NCBI HomoloGene Discover Homologs Help

Search HomoloGene for Co Clear

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 LOC479578 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

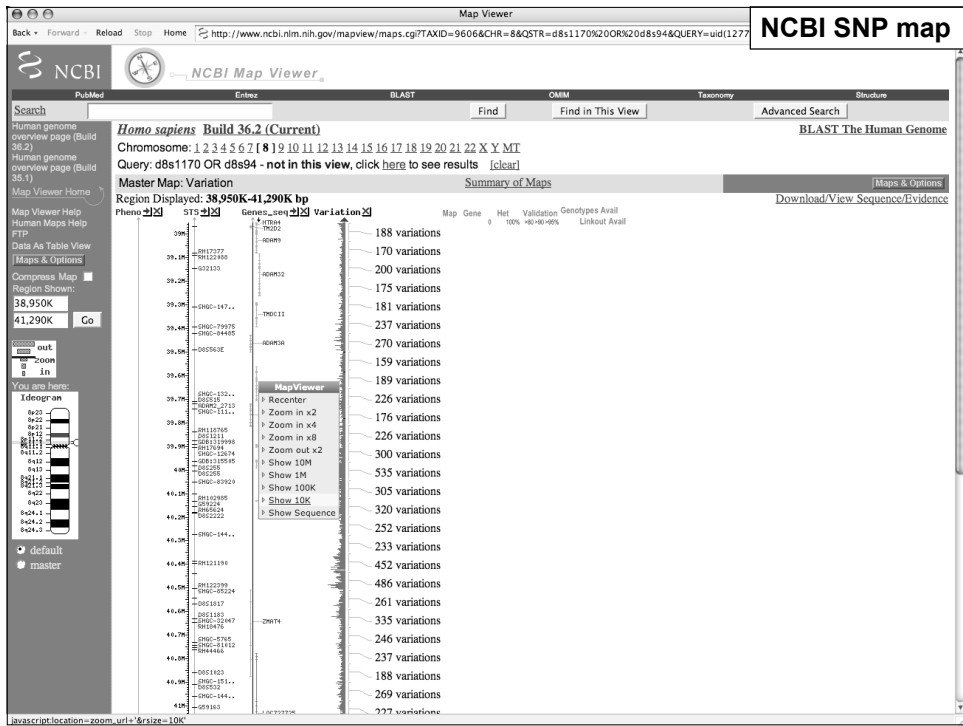
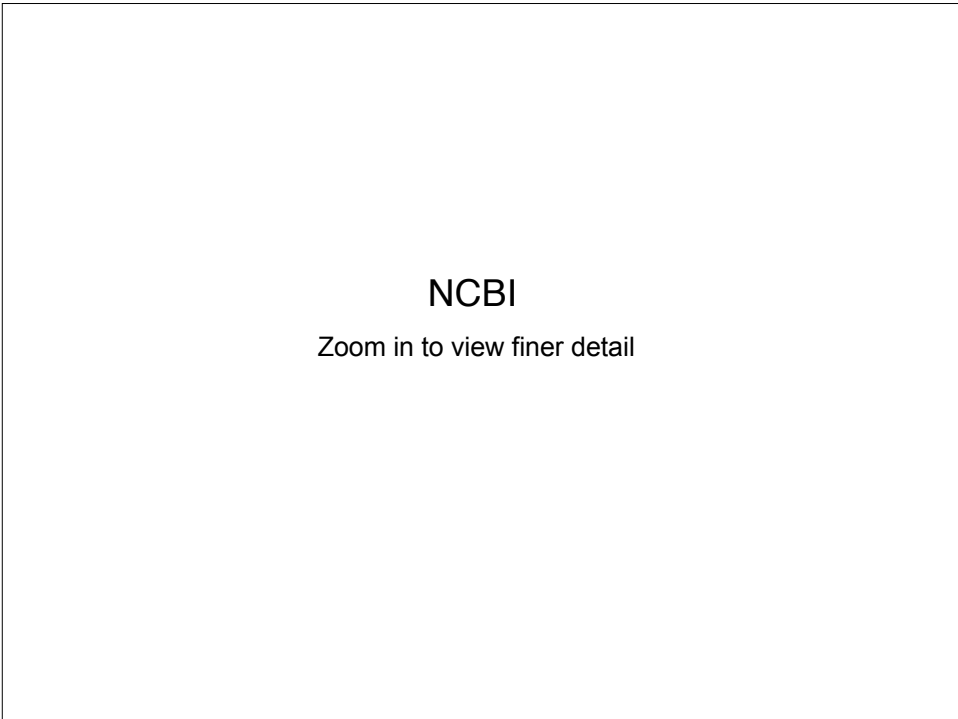
Conserved Domains
Conserved Domains from CDD found in protein sequences by rpsblast searching.

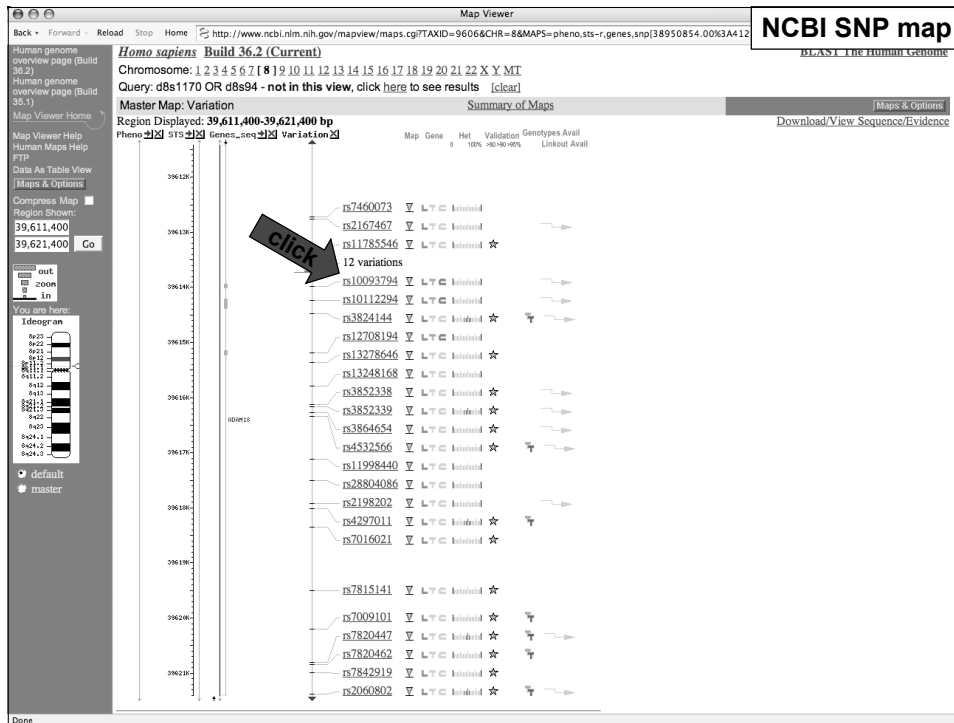
- smart00050
- DISIN, DISIN
- smart00638
- ACR, ACR
- pfam01562
- Pep_M12B_propep, Pep_M12B_propep
- pfam01421
- Reprolysin, Reprolysin

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.





NCBI Single Nucleotide Polymorphism **dbSNP**

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search Entrez SNP [] for [] Go

Reference SNP (refSNP)

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

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

```

tgacaaccac cttttcaact tttttcaatg tatttgaata tattttttt gttttttta
gataccacat ataagtlaga toatglaqa tttttttt ttgatoggo ttatttcaat
tagcaaatg tctctccatg tctctcctgt ttccacaaat gaaacagaaa ccatattcaa
ggcttaataa tatttccgca tgttgtat? Ttatatatat atatatatat atatatatat
atatatatat atatatatgt atatacaaaa tatataTCC ATAAAAATCT? GTctcatata
aaatatttga gaatttgttc aaattgtgac acattctctt atgtacttita aaacatttca
agattactta taatatctaa tacaatgctt aaatatactt teatlocat ggatccaana
taatactctc ATGGAATAA? AAATTTCTTT? TCCATATGGA? TTATATGOGA? TCTGAAATGA?
TGGCTGTAA? ACANAAAAAT?
&
TCCAGGTTAT TGGGCTGTCT AACACTGTAA GTTTTACTT? TTTCACATT? CCATTTTCAT
GAAAGTTTCT TTAATAATAA? TFGGTTCTCA TCTTGCCAA? TGAATAACT? AAATACATTT?
TATTTGTTTA ATCTGTAAAT? TACTGGAAAC? TGTTTTAT? TTTTATATG? ACCTTTTAA?
TGAATAACT? TGTTTCTT?
    
```

GeneView

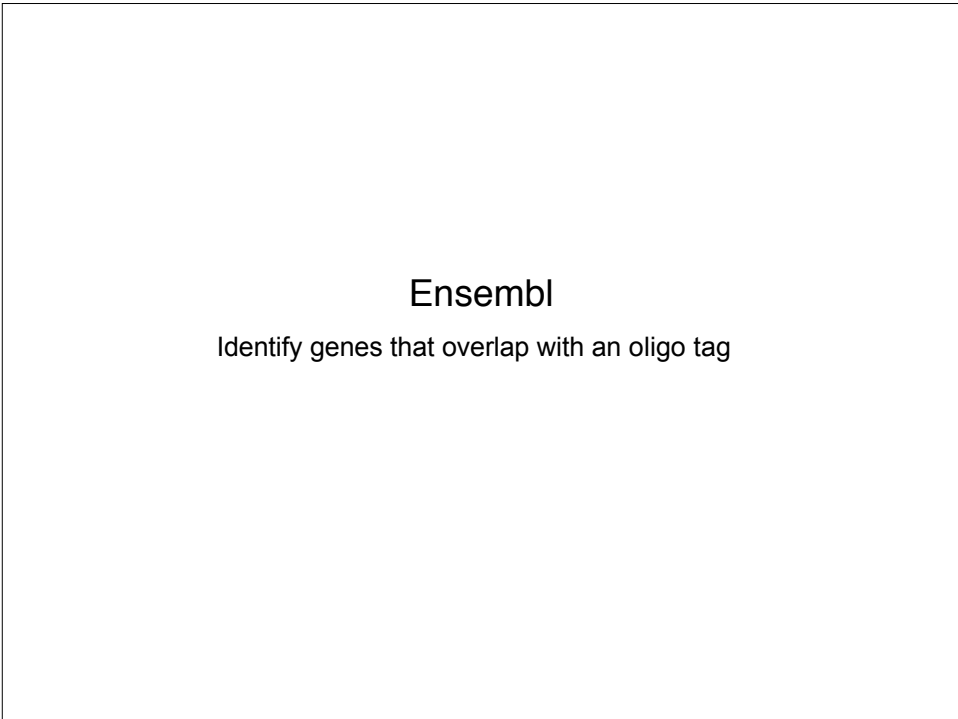
GeneView via analysis of contig annotation: ADAM18 a disintegrin and metalloproteinase domain 18
 Click to see [all] [cSNP] [has frequency] [double hit] [haplotype tagged] variations associated with this gene.

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	Amino acid pos
reference	NT_007995->NM_014237->NP_058558	9815211	forward	634	nonsynonymous	T	Phe [F]	I	212
Celera	NW_923907->NM_014237->NP_058558	9701448	forward	634	nonsynonymous	T	Phe [F]	I	212
					contig reference	G	Val [V]	I	212
					contig reference	G	Val [V]	I	212

NCBI
Find a chicken homolog of a human protein

The screenshot shows the NCBI BLAST search interface for chicken sequences. The browser address bar shows the URL: <http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9031>. The page title is "BLAST Chicken Sequences." The search input field contains the accession number "RP_001455". The database is set to "genome (reference only)" with 40216 sequences. The program is set to "tblastn: Compare a protein sequence against a nucleotide database". The optional parameters are: Expect: 0.01, Filter: low complexity, Descriptions: 100, and Alignments: 100. A black arrow points to the "Program:" dropdown menu with the word "select" written on it. The "Begin Search" button is visible at the bottom of the form.



Ensembl Genome Browser

Back Forward Reload Stop Home <http://www.ensembl.org/index.html> Search all Ensembl: Anything Go

Ensembl release 40 - Aug 2006 [Help](#)

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

New low-coverage genomes (*L. africana*, *Novemcinctus*, *E. telfairi*, *O. curvicolus*)

- Stickleback 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!** MPMU 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))
- Dasypus 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* (JGI 4.1)
- Danio rerio* (Zv6 | Vega)
- Takifugu rubripes* (FUGU 4.0)
- Tetraodon nigroviridis* (TETRAODON 7)
- Gasterosteus aculeatus* (**NEW!** BROAD S1)
- Oryzias latipes* (**Pre!** MEDAK 1)
- Ciona intestinalis* (JGI2)
- Ciona savignyi* (CSAV 2.0)
- Drosophila melanogaster* (**UPDATED!** EGCP 4)
- Anopheles gambiae* (AganP3)
- Aedes aegypti* (**NEW!** Aaeg. 1)
- Caenorhabditis elegans* (VS 150)
- Saccharomyces cerevisiae* (SCE 1)

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

Done

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

Chromosome 15

Chr. 15 band

DNA(contigs)

Markers

Ensembl Genes

nRNA Genes

EST Genes

Gene legend

© 2006 WTSI / 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

Detailed view

Chr. 15

Length

Forward strand

Reverse strand

Gene legend

© 2006 WTSI / 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 on Chromosome 15, specifically the 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=81. The interface is titled "Ensembl v40: Homo sapiens Features on Chromosome 15 54996168-55000187" and "Ensembl ContigView".

The main content area is divided into several sections:

- Chr. 15 Length:** A scale bar from 54,996,150 to 54,996,210 bp.
- Genscan:** Shows gene models for GENS CAN000004256 and GENS T0000001216.
- EST trans:** Shows TCF 12.
- Ensembl trans:** Shows NP_566161.1 and Ensembl Known Protein Coding.
- Blast hits:** Displays sequence alignments for various proteins, including Sth1321, Aha1, Bcl2, Bms5, Bsh4, Ecol1, Aha1, Ahr1, Ams8, Bms1, Bsa3, Bsk4, Cfr44, Eco18, and N388.

The left sidebar contains navigation and utility options:

- 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, Sitemap, Vega, Pre Ensembl, View previous release of page in Archivel, Stable Archivel link for this page, Archivel sites, Trace server.

Ensembl
Add features to the ContigView

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

Back Forward Reload Stop Home http://www.ensembl.org/Homo_sapiens/contigview?panel_bottom=on&h=15%3A54996168-55000187:h=BL **Ensembl ContigView**

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

Overview
Detailed view

select →

Features Comparative DAS Sources Repeats Decorations Export Image size Help

SNPs
Genotyped SNPs
Affy 100k SNP
Affy 500k SNP
IRNA
CpD islands
EpoLINE regions
First EF
Meths
Regulatory features
Vega assembly
Vega genes
Ensembl genes
ncRNA genes
EST genes
Targeted genome genes
GenScan
UniProtKB
UniProtKB (mammal)
UniProtKB (non-mammal)
Human proteins
Human RefSeqs
Other proteins
Uligene
EMBL mRNAs
MRNA
Fam
Human cDNAs
CDNAs
Sheep BAC ends
EST (ex.)
EST i
OLIGO HG_C110
OLIGO HG_F005
OLIGO HG_U133A
OLIGO HG_U133B
OLIGO HG_U133_2
OLIGO HG_U133_3
OLIGO HG_U133_4
OLIGO HG_U133_5
OLIGO HG_U133_6
OLIGO HG_U133_7
OLIGO HG_U133_8
OLIGO HG_U133_9
OLIGO HG_U133_10
OLIGO HG_U133_11
OLIGO HG_U133_12
OLIGO HG_U133_13
OLIGO HG_U133_14
OLIGO HG_U133_15
OLIGO HG_U133_16
OLIGO HG_U133_17
OLIGO HG_U133_18
OLIGO HG_U133_19
OLIGO HG_U133_20
OLIGO HG_U133_21
OLIGO HG_U133_22
OLIGO HG_U133_23
OLIGO HG_U133_24
OLIGO HG_U133_25
OLIGO HG_U133_26
OLIGO HG_U133_27
OLIGO HG_U133_28
OLIGO HG_U133_29
OLIGO HG_U133_30
OLIGO HG_U133_31
OLIGO HG_U133_32
OLIGO HG_U133_33
OLIGO HG_U133_34
OLIGO HG_U133_35
OLIGO HG_U133_36
OLIGO HG_U133_37
OLIGO HG_U133_38
OLIGO HG_U133_39
OLIGO HG_U133_40
OLIGO HG_U133_41
OLIGO HG_U133_42
OLIGO HG_U133_43
OLIGO HG_U133_44
OLIGO HG_U133_45
OLIGO HG_U133_46
OLIGO HG_U133_47
OLIGO HG_U133_48
OLIGO HG_U133_49
OLIGO HG_U133_50
OLIGO HG_U133_51
OLIGO HG_U133_52
OLIGO HG_U133_53
OLIGO HG_U133_54
OLIGO HG_U133_55
OLIGO HG_U133_56
OLIGO HG_U133_57
OLIGO HG_U133_58
OLIGO HG_U133_59
OLIGO HG_U133_60
OLIGO HG_U133_61
OLIGO HG_U133_62
OLIGO HG_U133_63
OLIGO HG_U133_64
OLIGO HG_U133_65
OLIGO HG_U133_66
OLIGO HG_U133_67
OLIGO HG_U133_68
OLIGO HG_U133_69
OLIGO HG_U133_70
OLIGO HG_U133_71
OLIGO HG_U133_72
OLIGO HG_U133_73
OLIGO HG_U133_74
OLIGO HG_U133_75
OLIGO HG_U133_76
OLIGO HG_U133_77
OLIGO HG_U133_78
OLIGO HG_U133_79
OLIGO HG_U133_80
OLIGO HG_U133_81
OLIGO HG_U133_82
OLIGO HG_U133_83
OLIGO HG_U133_84
OLIGO HG_U133_85
OLIGO HG_U133_86
OLIGO HG_U133_87
OLIGO HG_U133_88
OLIGO HG_U133_89
OLIGO HG_U133_90
OLIGO HG_U133_91
OLIGO HG_U133_92
OLIGO HG_U133_93
OLIGO HG_U133_94
OLIGO HG_U133_95
OLIGO HG_U133_96
OLIGO HG_U133_97
OLIGO HG_U133_98
OLIGO HG_U133_99
OLIGO HG_U133_100

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

Back Forward Reload Stop Home http://www.ensembl.org/Homo_sapiens/contigview?bottom=%7Cvariation%3Aon&w=4020&c=15%3A54996168-55000187 **Ensembl ContigView**

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

Basepair view

Features Comparative DAS Sources Repeats Decorations Export Image size Help

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

SNPs
Genotyped SNPs
Affy 100k SNP
Affy 500k SNP
IRNA
CpD islands
EpoLINE regions
First EF
Meths
Regulatory features
Vega assembly
Vega genes
Ensembl genes
ncRNA genes
EST genes
Targeted genome genes
GenScan
UniProtKB
UniProtKB (mammal)
UniProtKB (non-mammal)
Human proteins
Human RefSeqs
Other proteins
Uligene
EMBL mRNAs
MRNA
Fam
Human cDNAs
CDNAs
Sheep BAC ends
EST (ex.)
EST i
OLIGO HG_C110
OLIGO HG_F005
OLIGO HG_U133A
OLIGO HG_U133B
OLIGO HG_U133_2
OLIGO HG_U133_3
OLIGO HG_U133_4
OLIGO HG_U133_5
OLIGO HG_U133_6
OLIGO HG_U133_7
OLIGO HG_U133_8
OLIGO HG_U133_9
OLIGO HG_U133_10
OLIGO HG_U133_11
OLIGO HG_U133_12
OLIGO HG_U133_13
OLIGO HG_U133_14
OLIGO HG_U133_15
OLIGO HG_U133_16
OLIGO HG_U133_17
OLIGO HG_U133_18
OLIGO HG_U133_19
OLIGO HG_U133_20
OLIGO HG_U133_21
OLIGO HG_U133_22
OLIGO HG_U133_23
OLIGO HG_U133_24
OLIGO HG_U133_25
OLIGO HG_U133_26
OLIGO HG_U133_27
OLIGO HG_U133_28
OLIGO HG_U133_29
OLIGO HG_U133_30
OLIGO HG_U133_31
OLIGO HG_U133_32
OLIGO HG_U133_33
OLIGO HG_U133_34
OLIGO HG_U133_35
OLIGO HG_U133_36
OLIGO HG_U133_37
OLIGO HG_U133_38
OLIGO HG_U133_39
OLIGO HG_U133_40
OLIGO HG_U133_41
OLIGO HG_U133_42
OLIGO HG_U133_43
OLIGO HG_U133_44
OLIGO HG_U133_45
OLIGO HG_U133_46
OLIGO HG_U133_47
OLIGO HG_U133_48
OLIGO HG_U133_49
OLIGO HG_U133_50
OLIGO HG_U133_51
OLIGO HG_U133_52
OLIGO HG_U133_53
OLIGO HG_U133_54
OLIGO HG_U133_55
OLIGO HG_U133_56
OLIGO HG_U133_57
OLIGO HG_U133_58
OLIGO HG_U133_59
OLIGO HG_U133_60
OLIGO HG_U133_61
OLIGO HG_U133_62
OLIGO HG_U133_63
OLIGO HG_U133_64
OLIGO HG_U133_65
OLIGO HG_U133_66
OLIGO HG_U133_67
OLIGO HG_U133_68
OLIGO HG_U133_69
OLIGO HG_U133_70
OLIGO HG_U133_71
OLIGO HG_U133_72
OLIGO HG_U133_73
OLIGO HG_U133_74
OLIGO HG_U133_75
OLIGO HG_U133_76
OLIGO HG_U133_77
OLIGO HG_U133_78
OLIGO HG_U133_79
OLIGO HG_U133_80
OLIGO HG_U133_81
OLIGO HG_U133_82
OLIGO HG_U133_83
OLIGO HG_U133_84
OLIGO HG_U133_85
OLIGO HG_U133_86
OLIGO HG_U133_87
OLIGO HG_U133_88
OLIGO HG_U133_89
OLIGO HG_U133_90
OLIGO HG_U133_91
OLIGO HG_U133_92
OLIGO HG_U133_93
OLIGO HG_U133_94
OLIGO HG_U133_95
OLIGO HG_U133_96
OLIGO HG_U133_97
OLIGO HG_U133_98
OLIGO HG_U133_99
OLIGO HG_U133_100

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

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

select →

Pre Ensembl
View previous release of page | Release 39: Jun 2006
Stable Release 38: Apr 2006
Archiv Release 37: Feb 2006
Trace Release 36: Dec 2005
Release 35: Nov 2005
Release 34: Oct 2005

The screenshot displays the Ensembl v37 Human ContigView interface. The browser address bar shows the URL: http://feb2006.archive.ensembl.org/Homo_sapiens/contigview?panel_top=off;15k3AS4996168-55000187;h=BLAS. The page title is "Ensembl v37: Homo sapiens Features on Chromosome 15 54996168-55000187". The main content area shows a genomic track for Chromosome 15, with coordinates ranging from 54,996,168 to 55,000,187. The track displays various genomic features, including CDNs, Human cDNAs, EMBL mRNAs, UniGene, Human RefSeqs, Human proteins, GenScan, EST trans., Ensembl trans., Blast hits, DNase (contigs), and EST trans. The interface includes navigation controls such as "Jump to region", "Zoom", and "Refresh". A sidebar on the left provides navigation options like "View of Chromosome 15", "Graphical overview", and "Export information about region".

Ensembl

Get additional information about the gene, transcripts, and exons

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

Ensembl Human ContigView

Search of Human: Anything

e.g. AL138722.15.1.44776, AL355340.17.1.112442

Ensembl release 40 - Aug 2006

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

Chromosome 15
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
Sitemap
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
SNP
Length

Gene legend
EST gene
Ensembl Known Protein Coding
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 v40: Homo sapiens Gene report for ENSG00000140262

Ensembl Human GeneView

Search of Human: Anything

e.g. ENSG00000139618, ENSG00000128573

Ensembl release 40 - Aug 2006

ENSG00000140262

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	Gene	Transcript info	Exon info	Peptide info
ENST00000267811	TCF12	[Transcript info]	[Exon info]	[Peptide info]
ENST00000333725	NP_996919.1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000343827	NP_996923.1	[Transcript info]	[Exon info]	[Peptide info]

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

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
AC010999.6.1.196602

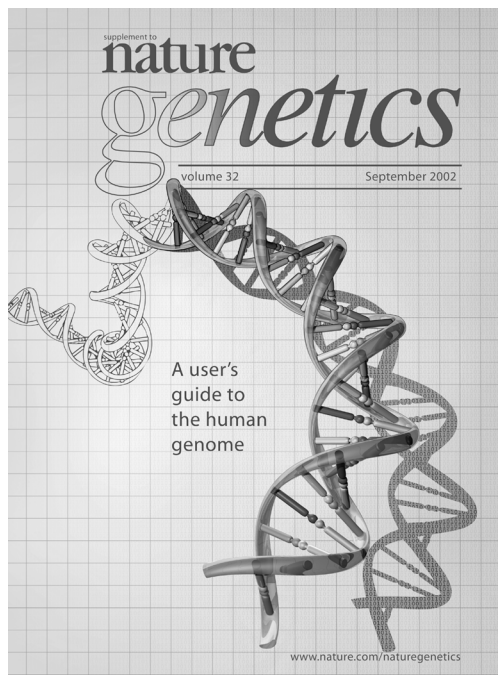
Orthologue Prediction

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

Species	Type	Gene identifier
<i>Clona savignyi</i>	1 to many	ENSCSAVG00000011705 (Novel Ensembl prediction) [Info] [Align] No description
<i>Canis familiaris</i>	1 to 1	ENSCAFGG00000016200 (TCF12) [Info] [Align] PROTID: similar to Transcription factor 12 isoform 9 (Ensembl) [Info] [Align] [Peptide] [DP_535491]
<i>Aedes aegypti</i>	1 to many	AAELI010226 (Novel Ensembl prediction) [Info] [Align] duphtr12
<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	ENSMUSG00000032228 (TCF12) [Info] [Align] transcription factor 12 [Source: MarkerSymbol, Acc: #901101877]
<i>Pan troglodytes</i>	1 to 1	ENSPTRG00000007109 (TCF12) [Info] [Align] No description

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.