

NATIONAL HUMAN GENOME RESEARCH INSTITUTE *Division of Intramural Research*



*Current Topics in Genome Analysis
Spring 2008*

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>

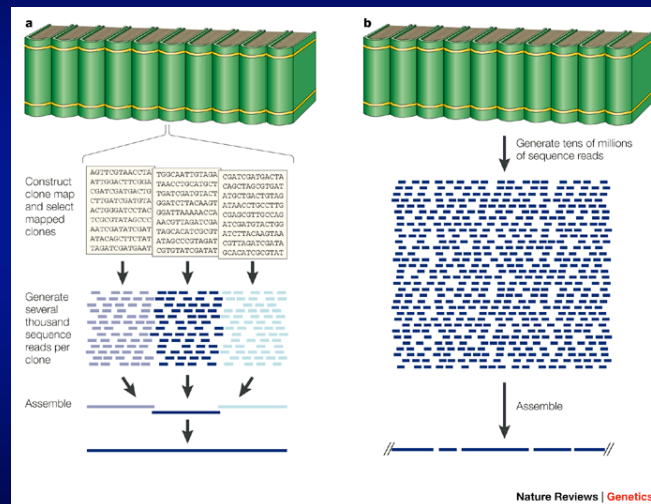
Types of data integrated in genome browsers

- Same starting material for all genome browsers: genomic sequence
- Annotations calculated independently by each genome browser
 - Genes
 - RefSeq mRNAs (non-redundant)
 - GenBank mRNAs (redundant)
 - 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 archive of all genome assemblies and annotations; NCBI provides only limited archive
- IF YOU ARE COMPARING DATA FROM DIFFERENT GENOME BROWSERS, MAKE SURE YOU ARE LOOKING AT THE SAME VERSION OF THE ASSEMBLY

Genome Assembly Versions

	Same assembly?	UCSC	NCBI	Ensembl
Human	Yes	Mar 2006/hg18/Build 36.1	Build 36.2	Build 36
Mouse	YES	July 2007/mm9/Build 37	Build 37.1	Build 37
Dog		May 2005 /canFam 2.0	Build 2.1/CanFam 2.1	CanFam 2.0
Zebrafish	NO	July 2007/danRer5/Zv7	Zv6/build 2.1	Zv7
Rhesus	YES	Jan 2006/rheMac2/v.1.0, Mmul_051212	Build 1.1/v.1.0, Mmul_051212	Mmul_1

NCBI Reference Sequences (RefSeqs)

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

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

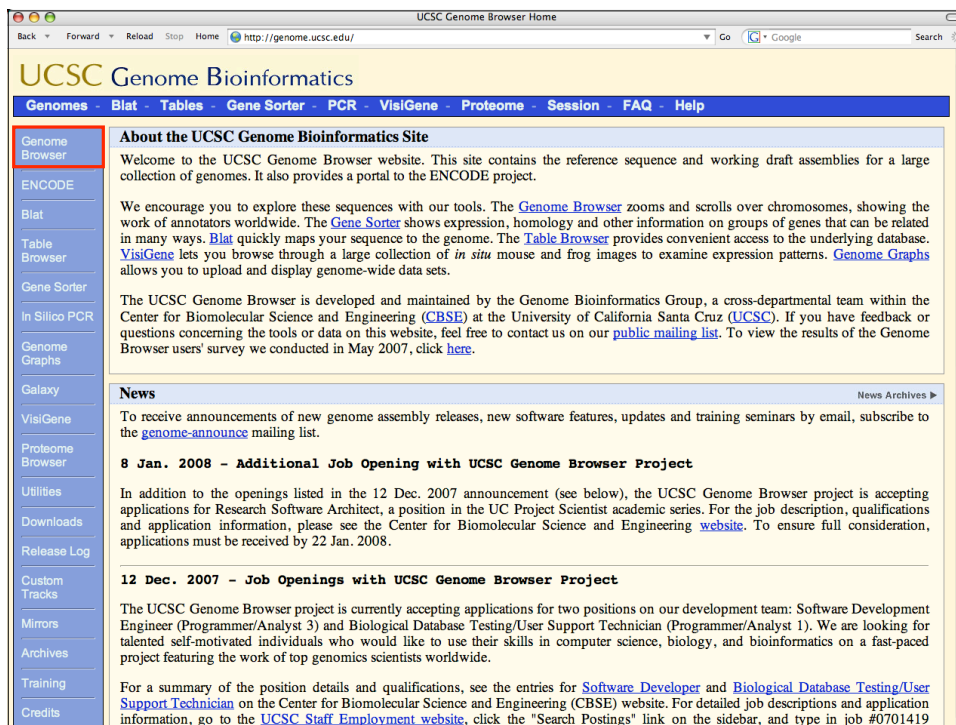
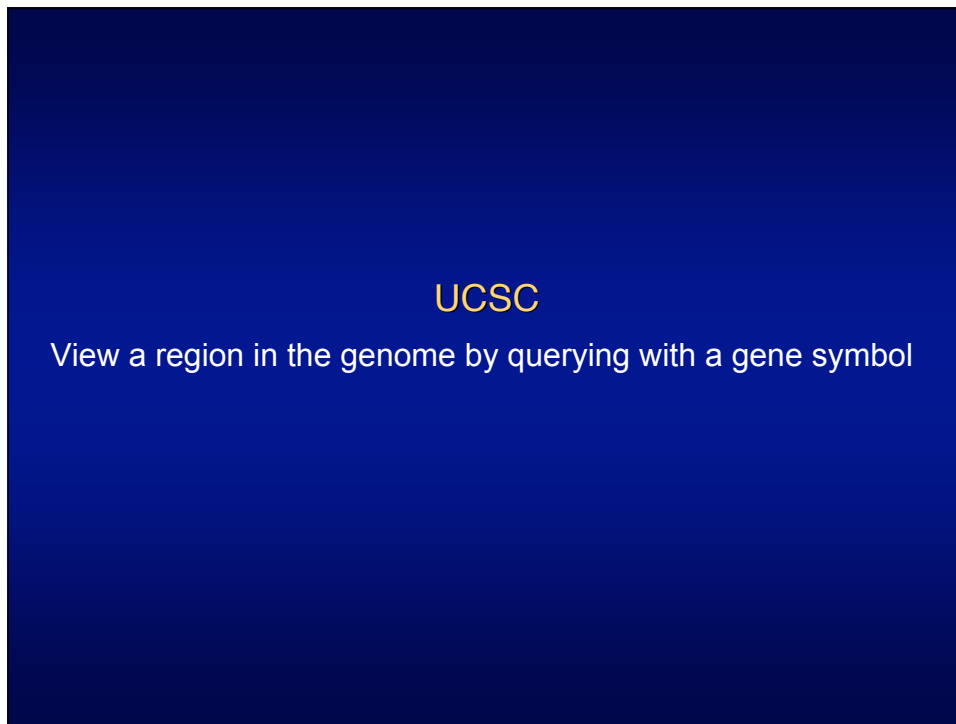
<http://www.ncbi.nlm.nih.gov/RefSeq/key.html>

```

LOCUS       NM_001101                1793 bp    mRNA    linear   PRI 14-JAN-2008
DEFINITION Homo sapiens actin, beta (ACTB), mRNA.
ACCESSION  NM_001101
VERSION    NM_001101.2  GI:5016088
KEYWORDS   .
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Theria; Euarchontoglires; Primates; Haplorhini;
          Villebeck,L., Moparhi,S.B., Lindgren,M., Hammarstrom,P. and
          Jonsson,B.H.
TITLE      Domain-specific chaperone-induced expansion is required for
          beta-actin folding; a comparison of beta-actin conformations upon
          interactions with GroEL and tail-less complex polypeptide 1 ring
          complex (TRiC)
JOURNAL    Biochemistry 46 (44), 12639-12647 (2007)
PUBMED    17333480
COMMENT    REVIEWED REFSEQ: This record has been curated by NCBI staff. The
          reference sequence was derived from X00351.1 and X63432.1.
          On Jun 8, 1999 this sequence version replaced gi:450185.
          Summary: This gene encodes one of six different actin proteins.
          Actins are highly conserved proteins that are involved in cell
          motility, structure, and integrity. This actin is a major
          constituent of the contractile apparatus and one of the two
          nonmuscle cytoskeletal actins.
          Publication Note: This RefSeq record includes a subset of the
          publications that are available for this gene. Please see the
          Entrez Gene record to access additional publications.
          COMPLETENESS: complete on the 3' end.
          ..
CDS         74..1201
          /gene="ACTB"
          /locus_tag="actin filament; cytoskeleton; T1P60 histone
          acetyltransferase complex [PMID 10956108]"
          /GO_function="ATP binding; nucleotide binding; protein
          binding [PMID 15527767]; structural constituent of
          cytoskeleton"
          /note="beta cytoskeletal actin; PS1TP5-binding protein 1;
          actin, cytoplasmic 1; beta-actin"
          /codon_start=1
          /product="beta actin"
          /protein_id="NP_001092.1"
          /db_xref="GI:450185"
          /db_xref="CCDS:CCDS5341.1"
          /db_xref="GeneID:61"
          /db_xref="HGNC:137"
          /db_xref="HPRD:00012"
          /db_xref="MIM:103610"
          /translation="MDDDIALLVVDNGSOMCKAGFADDAPRAVFPFSIVGRPHIQGVN
          VMGQKDSYVGDASGRKGLTLKYPFIEGIVNNDMEKIMHFFYNELRVAPENP
          VLTETAPLNFANRKNQYIMFTFPPANVVAIDVLSLYAGSTTIVNDSGDVY
          HTPVIVSVALPHAILRLDLACRDLTDYIMKILTERGYSFTTAREIIVRDIKELCY
          VALDFQENATASSSLSKSYELPDQGVITIGNRFKCPALFPSPFLGMSGCCHE
          FTNSIHSQVPISSGLANFVSGSTWTFYADRMKSTPLASFTWIKIAPPE
          RKYVWVGSLASLSTFQGMWISKQYDESQSPVIRKCF"
          ..
ORIGIN      1  cggctcgcgc  ccgcgagcac  agagctctgc  ctttgcgatg  cgcgcgccgc  tccacaccgc
          61  ccgcacgctc  accatgatgt  atgatctcgc  cgcgctctgc  gtcgacaacg  gctccggcat
          121 gtcgaagcgc  ggtctcgcgc  gcgacgatgc  cccccggccc  gctctcccc  ccatgctggg

```

Beta actin mRNA RefSeq



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: 820

Request:

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
RH18061;RH80175	Displays region between STS markers RH18061;RH80175. 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

UCSC Genes

ADAM2 (uc003kx1.1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein

ADAM2 (uc003kx2.1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein

ADAM2 (uc003kxj.1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein

ADAM2B (uc003kdy.1) at chr8:24207525-24268671 - ADAM metallopeptidase domain 28 isoform 1

ADAM28 (uc003kdx.1) at chr8:24207525-24268671 - ADAM metallopeptidase domain 28 isoform 3

ADAM22 (uc003ajp.1) at chr7:87401638-8764385 - ADAM metallopeptidase domain 22 isoform 4

ADAM22 (uc003ajo.1) at chr7:87401638-8764385 - ADAM metallopeptidase domain 22 isoform 3

ADAM22 (uc003ajn.1) at chr7:87401638-8764385 - ADAM metallopeptidase domain 22 isoform 1

ADAM22 (uc003ajm.1) at chr7:87401638-8764385 - ADAM metallopeptidase domain 22 isoform 2

ADAM22 (uc003ajk.1) at chr7:87401638-8764385 - ADAM metallopeptidase domain 22 isoform 5

ADAM22 (uc003ajl.1) at chr7:87401638-8764385 - ADAM metallopeptidase domain 22 isoform 4

ADAM22 (uc003ajh.1) at chr7:87401638-8764385 - ADAM metallopeptidase domain 22 isoform 5

ADAM29 (uc003iue.1) at chr4:17607634-176135906 - ADAM metallopeptidase domain 29 preproprotein

ADAM29 (uc003iud.1) at chr4:17607634-176135906 - ADAM metallopeptidase domain 29 preproprotein

ADAM29 (uc003iug.1) at chr4:17607634-176135906 - ADAM metallopeptidase domain 29 preproprotein

ADAM23 (uc002vbg.1) at chr2:207016613-207190924 - ADAM metallopeptidase domain 23 preproprotein

ADAM20 (uc001xme.1) at chr14:70058831-70071485 - ADAM metallopeptidase domain 20 preproprotein

ADAM21 (uc001xmd.1) at chr14:69993970-69996375 - ADAM metallopeptidase domain 21 preproprotein

RefSeq Genes

ADAM2 at chr8:39720412-39814936 - (NM_001464) ADAM metallopeptidase domain 2 proprotein

ADAM20 at chr14:70058832-70071485 - (NM_003814) ADAM metallopeptidase domain 20 preproprotein

ADAM21 at chr14:69993970-69996374 - (NM_003813) ADAM metallopeptidase domain 21 preproprotein

ADAM22 at chr7:87401638-8764385 - (NM_004194) ADAM metallopeptidase domain 22 isoform 4

ADAM22 at chr7:87401638-8764384 - (NM_021721) ADAM metallopeptidase domain 22 isoform 5

ADAM22 at chr7:87401638-8764383 - (NM_021722) ADAM metallopeptidase domain 22 isoform 2

ADAM22 at chr7:87401638-8764383 - (NM_021723) ADAM metallopeptidase domain 22 isoform 1

ADAM22 at chr7:87401638-8764383 - (NM_016351) ADAM metallopeptidase domain 22 isoform 3

ADAM23 at chr2:207016613-207190922 - (NM_003812) ADAM metallopeptidase domain 23 preproprotein

ADAM28 at chr8:24207525-24268670 - (NM_021777) ADAM metallopeptidase domain 28 isoform 3

ADAM28 at chr8:24207525-24268670 - (NM_014265) ADAM metallopeptidase domain 28 isoform 1

ADAM29 at chr4:17607634-176135906 - (NM_014269) ADAM metallopeptidase domain 29 preproprotein

Non-Human RefSeq Genes

ADAM2 at chr8:39723140-39814873 - (NM_001082677) ADAM metallopeptidase domain 2 (fertilin beta)

ADAM2 at chr8:39723146-39814902 - (NM_219571) ADAM metallopeptidase domain 2

ADAM2 at chr8:39721536-39813888 - (NM_174228) ADAM metallopeptidase domain 2

Adam2 at chr8:39723147-39801569 - (NM_009618) a disintegrin and metalloprotease domain 2

ADAM2 at chr8:39723147-39813877 - (NM_020077) a disintegrin and metalloprotease domain 2

Adam21 at chr14:69998947-69996354 - (NM_020300) a disintegrin and metallopeptidase domain 21

Adam21 at chr14:69781958-70061197 - (NM_020330) a disintegrin and metallopeptidase domain 21

Adam22 at chr7:87401747-87649282 - (NM_001007220) a disintegrin and metalloprotease domain 22

Adam22 at chr7:87401747-87649282 - (NM_001007221) a disintegrin and metalloprotease domain 22

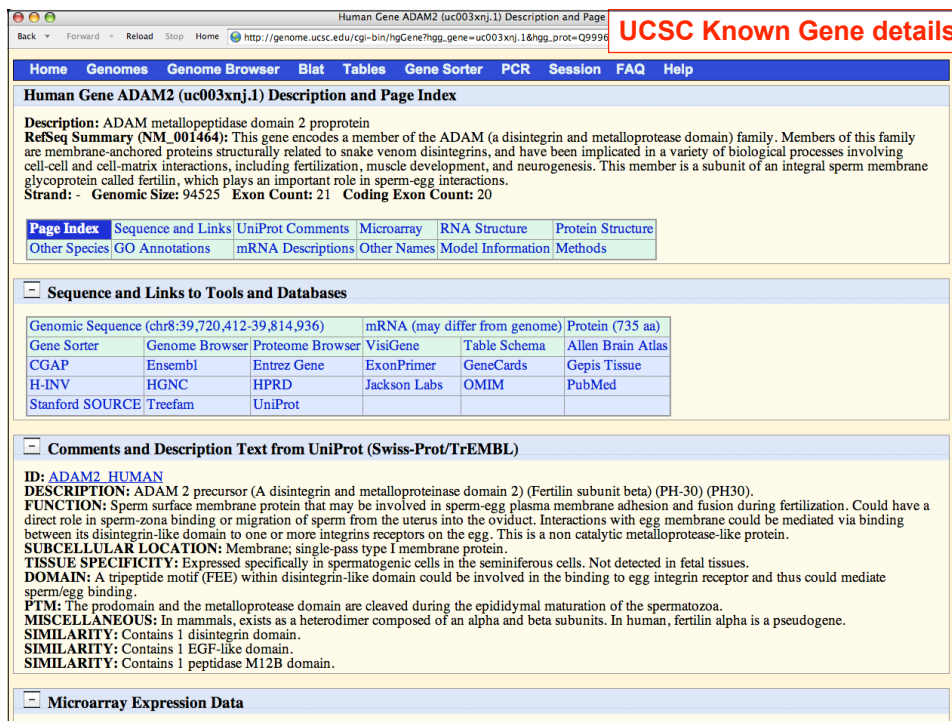
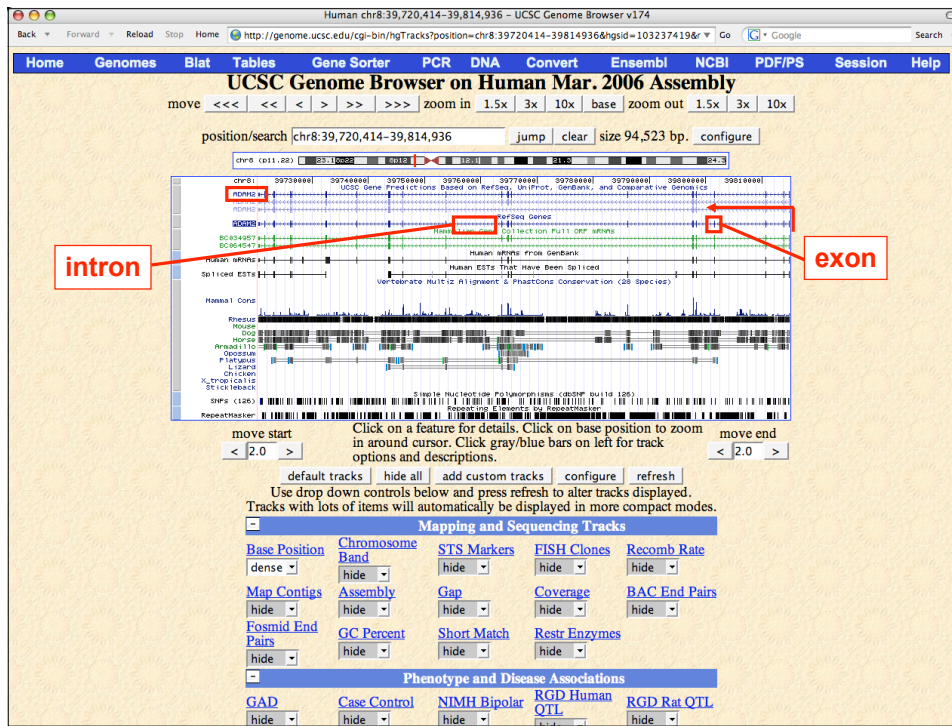
Adam22 at chr7:87401747-87649282 - (NM_001098225) a disintegrin and metalloprotease domain 22

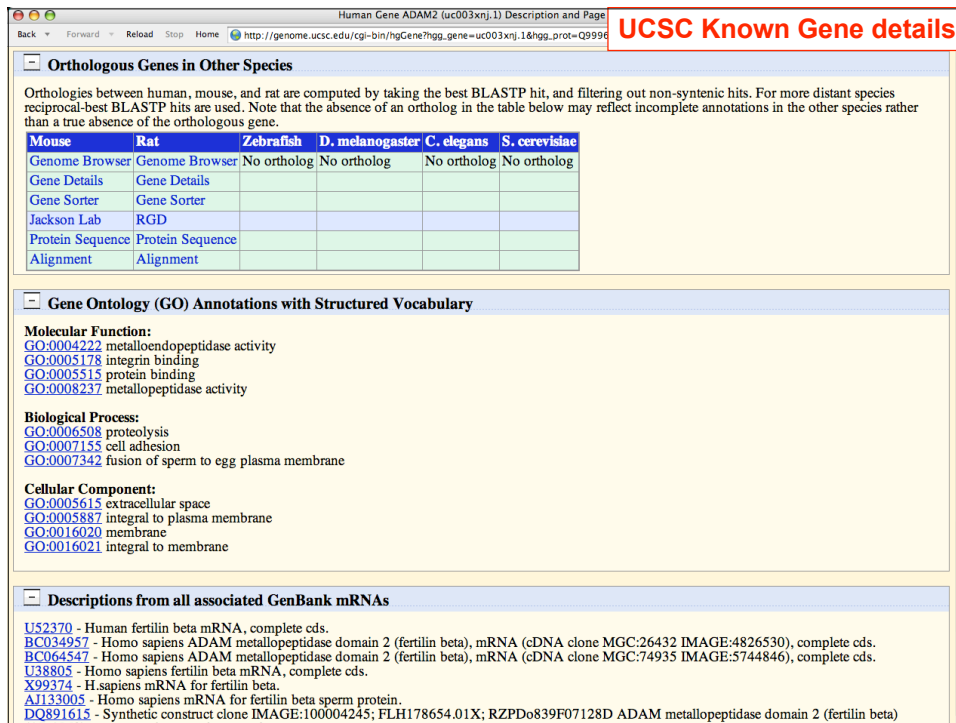
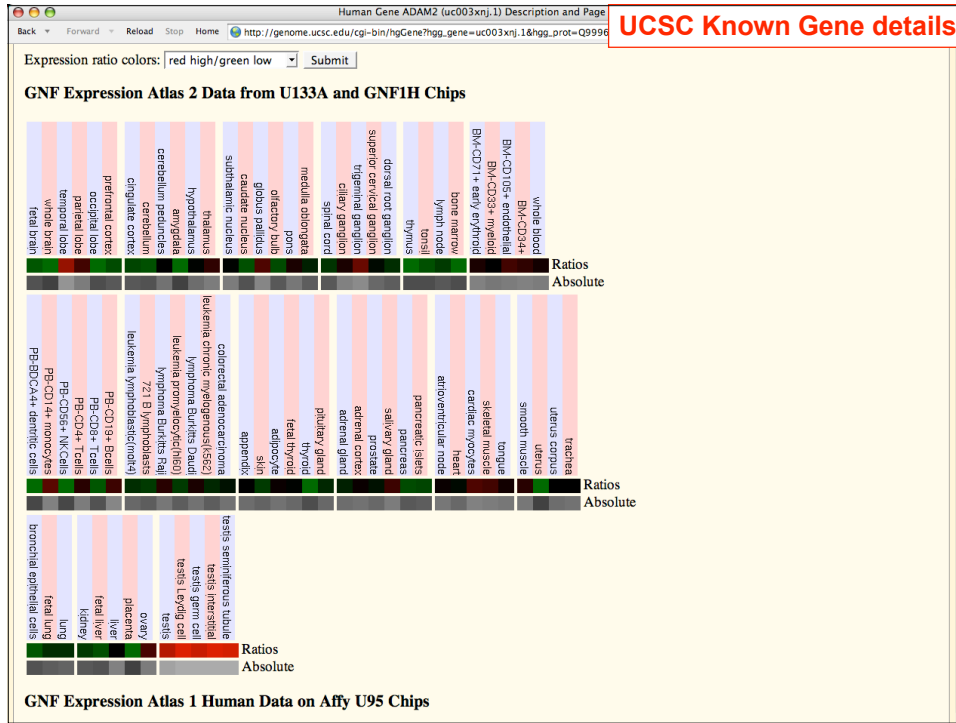
Adam23 at chr2:207016503-207194450 - (NM_011780) a disintegrin and metallopeptidase domain 23

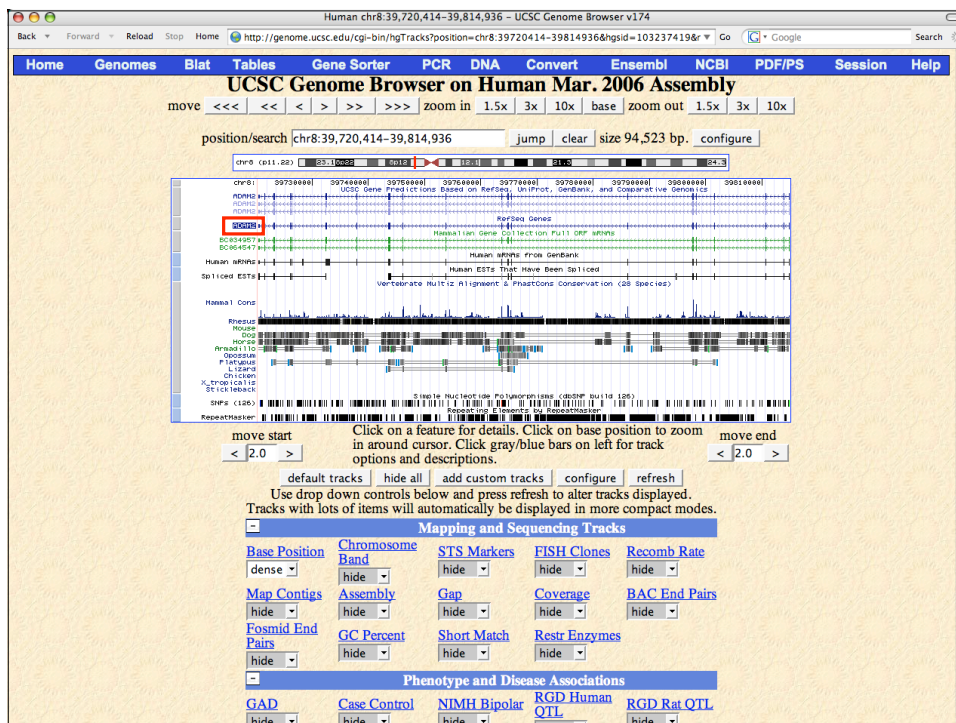
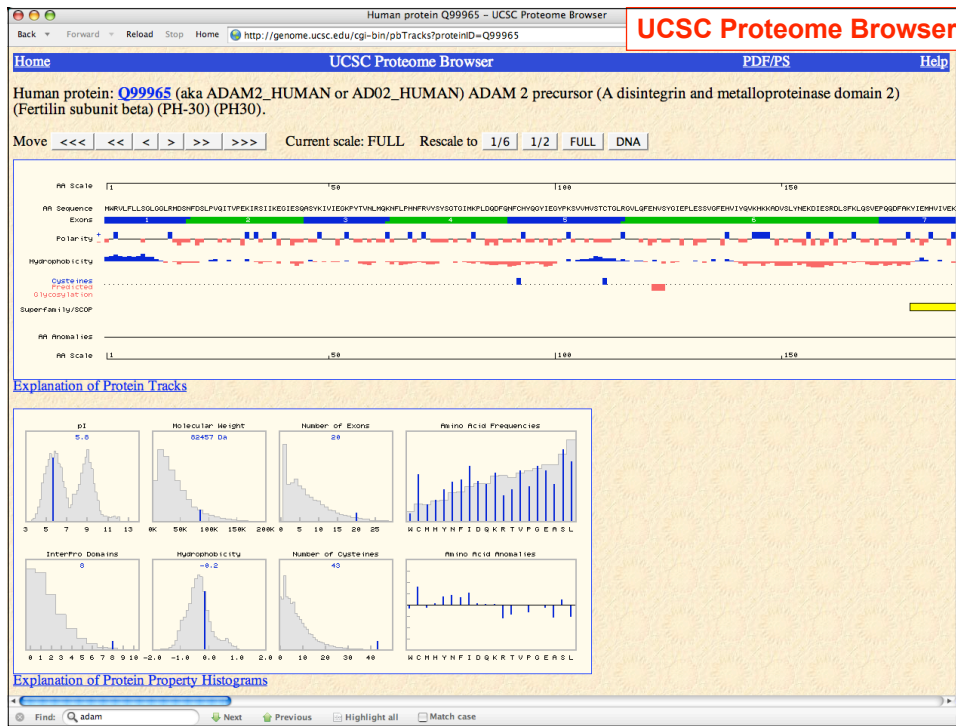
Adam23 at chr2:207016503-207194450 - (NM_001029899) a disintegrin and metalloprotease domain 23

Adam24 at chr1:18904672-188906153 - (NM_010086) a disintegrin and metalloprotease domain 24

Adam25 at chr1:18904411-188906248 - (NM_011781) a disintegrin and metalloprotease domain 25







UCSC RefSeq Gene details

RefSeq: [NM_001464.3](#) Status: **Reviewed**
 CCDS: [CCDS34884.1](#)
 CDS: 3' complete
 OMIM: [601533](#)
 Entrez Gene: [2515](#)
 PubMed on Gene: [ADAM2](#)
 PubMed on Product: [ADAM 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

[View details of parts of alignment within browser window.](#)

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

UCSC RefSeq Gene details

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session 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

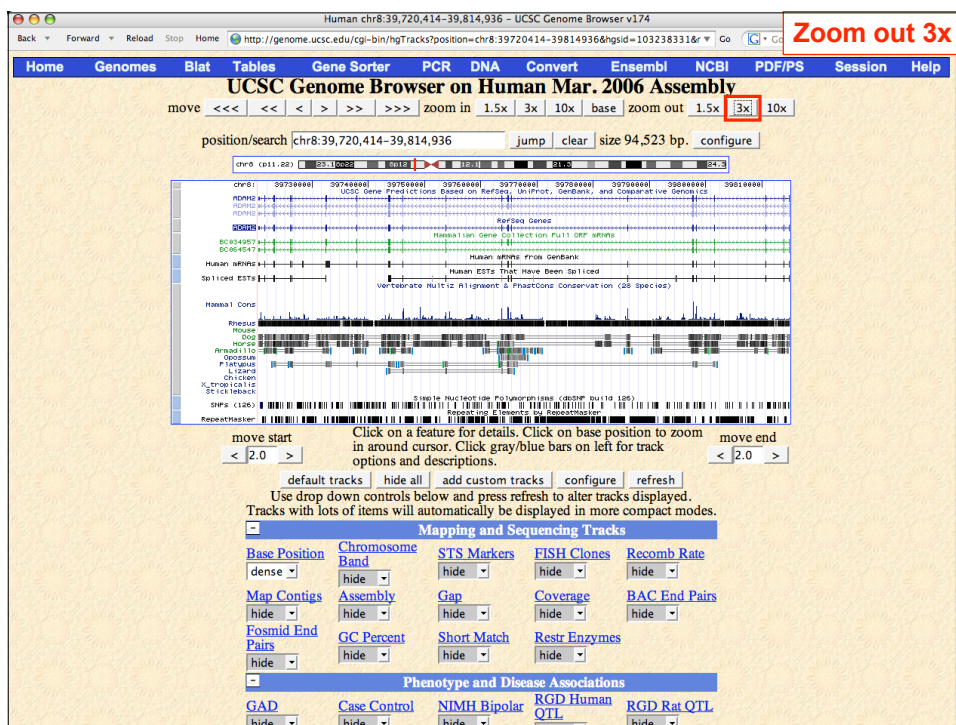
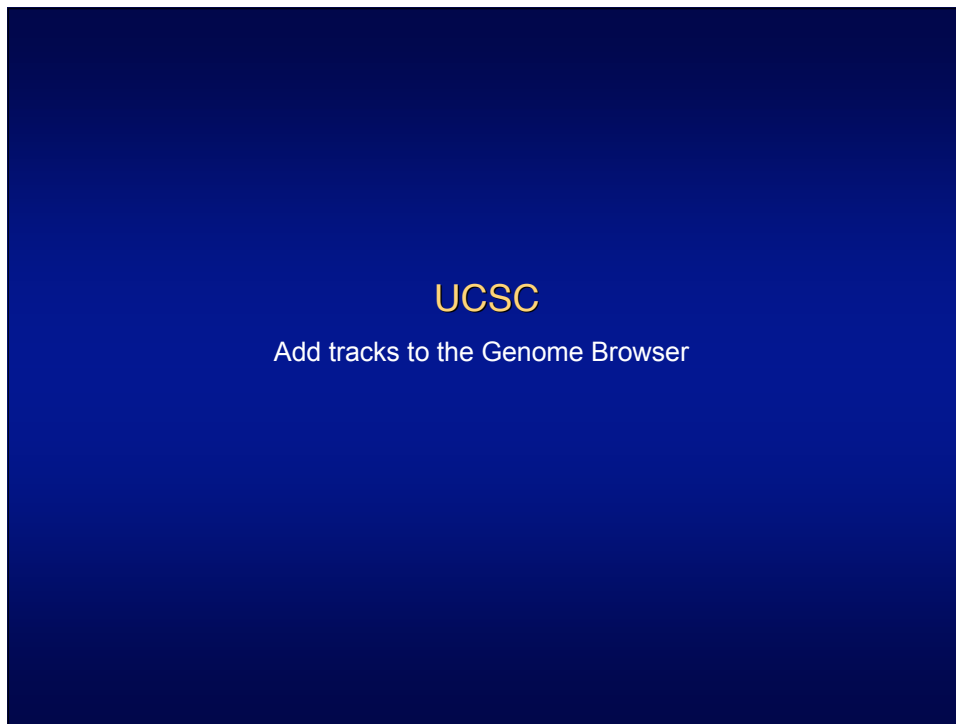
Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

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
atatacaacgtagaaacccctacttatcacocaaagcctgtagaaga
atgttatagctaaatatttttaaatagctggaacataaacaacaa
aatatctaacagtaaaaatggaacacaagaagtggttatatttaata
attgtatatacaacaaatgaggttaaacagaactatgcttggataga
accttaacatcatctatataaagaacccacacatgaaagatgagttg
gattgcttactctgcaaaaagtcaaaaacagacaacaaacgaatcttgg
ttgtttagaagtcagtggttgaggtggaatctgggatttgggtgggt
ctttttcaatctttcaactgggtaactagtttaaggttttttttcaac
ttgaatataatgacctggaactatgattattataacttttttctc
gttttttggctctcttttttttttttttttttttttttttttttttga
ttcgtctctcaacacagctggagtgcaaggttcaaggtctctctgcaac
tgcaacctgctccttaggttcaagcagattctctgctcagctctccg
agttagctgggatttccaggaacccgcaacatgctgctaatttttttt
gtatttttagtaacaggggtttcaocatttggcgaacatggtctcga
actcctgatcagtggtttatataatcaattgaaattttacttaagaagt
gtttataaatctctctgttctcagctgttgaagtgatttggttgttgc
tgttgccatttaataggatcaagctcccgtaggtgcttggctctcgaag
acagggctcagagctgcaagcttcaacagcgcacacccaacacactcag
cccacctgggctctcccaagcctacctcttccaggtgctggccggg
```



Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v174

http://genome.ucsc.edu/cgi-bin/hgTracks?hgid=103238331&hgout2=+3x+&position=chr8%3A397,204,14-39,814,936

UCSC Gene Predictions Based on RefSeq, UniProt, GenBank, and Comparative Genomics

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

Vertebrate Multiz Alignment & PhastCons Conservation (20 Species)

RepeatMasker

move start < 2.0 > move end < 2.0 >

Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

default tracks hide all add custom tracks configure refresh

click

Expression and Regulation

Affy All Exon	Affy HuEx 1.0	GNF Atlas 2	Allen Brain	GNF Ratio
hide	hide	hide	hide	hide
Bertone Yale TAR	Affy U133	Affy GNF1H	Affy U133Plus2	Affy U95
hide	hide	hide	hide	hide
CpG Islands	FirstEF	SwitchGear TSS	Eponine TSS	Vista Enhancers
hide	hide	hide	hide	hide
TFBS Conserved	GIS PET...	LIUCSD TAF1...	ORegAnno	Affy Txn...
hide	hide	hide	hide	hide
hide dense squish pack full	7X Reg Potential	Uppsala ChIP...		
	hide	hide		

Comparative Genomics

Human chr8:39,625,891-39,909,459 - UCSC Genome Browser v174

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

UCSC TFBS Track

Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session 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 chr8:39,625,891-39,909,459 jump clear size 283,569 bp. configure

chr8 (31,28-411,21)

UCSC Gene Predictions Based on RefSeq, UniProt, GenBank, and Comparative Genomics

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

HMRF Conserved Transcription Factor Binding Sites

Vertebrate Multiz Alignment & PhastCons Conservation (20 Species)

RepeatMasker

move start < 2.0 > move end < 2.0 >

Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

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

click x 3

UCSC Genome Browser on Human Mar. 2006 Assembly

position/search chr8:39,884,779-39,895,280

click

UTR CDS

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
Fosmid End Pairs	GC Percent	Short Match	Restr Enzymes	
hide	hide	hide	hide	

Phenotype and Disease Associations

GAD	Case Control	NIMH Bipolar	RGD Human QTL	RGD Rat QTL
-----	--------------	--------------	---------------	-------------

UCSC TFBS Track details

HMR Conserved Transcription Factor Binding Sites (VSIRF2_01)

Transcription Factor Binding Site information:

Item: VSIRF2_01
 Transfac matrix link: [M00063](#)
 Score: 891
 zScore: 3.04
 Strand: -
 Position: [chr8:39889441-39889453](#)
 Band: [8p11.22](#)
 Genomic Size: 13
[View DNA for this feature](#)

Transcription Factors known to bind to this site:

Factor: IRF-2
 Species: mouse
 SwissProt ID: P23906
[View table schema](#)

Data last updated: 2007-07-17

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 (v7.0) 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 putative 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.



Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v174

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

chr8 (p11.22) 39730000 39740000 39750000 39760000 39770000 39780000 39790000 39800000 39810000

RefSeq Genes
HAWAIIAN GENE COLLECTION FULL SNP ARRAY
Human mRNAs From GenBank
Human ESTs That Have Been Spliced
HMR Conserved Transcription Factor Binding Sites
Vertebrate Multiple Alignment & PhastCons Conservation (28 Species)
Hawaii Cons
Rhesus
Mouse
Dog
Monkey
Pig
Opossum
Platypus
Chicken
X_TropLoc1112
ESTs (BRUCE)
SNPs (126)
RepeatMasker

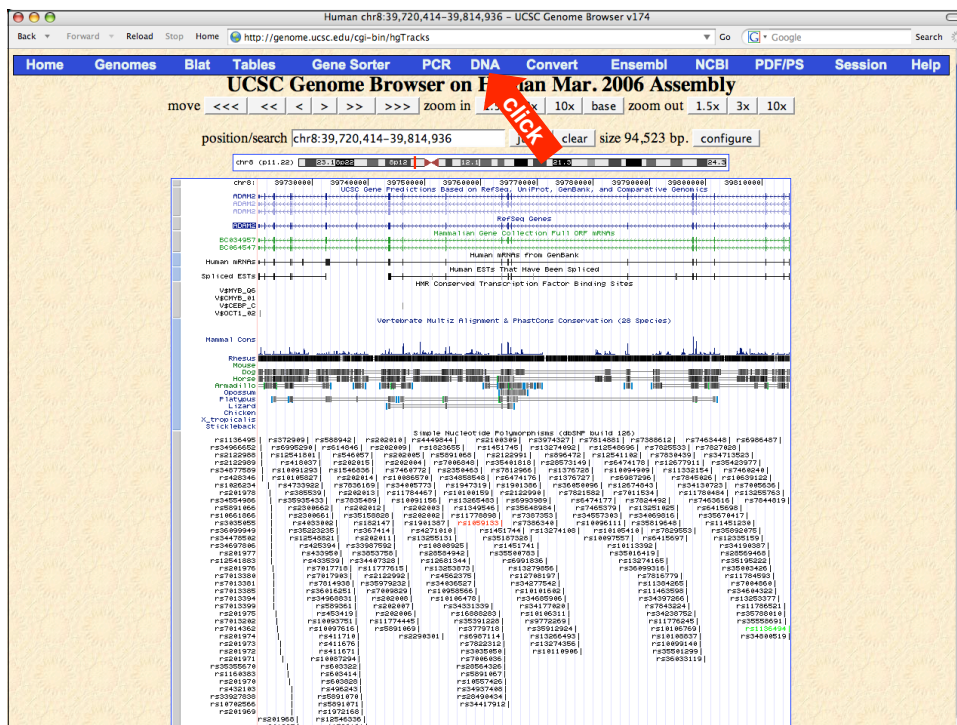
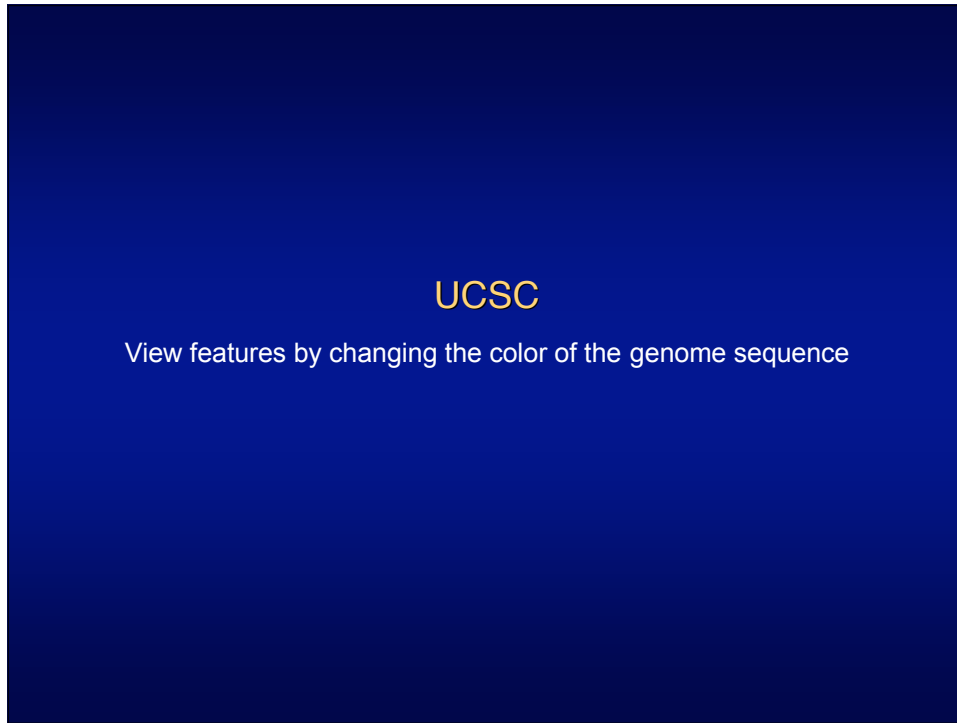
move start < 2.0 > Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions. move end < 2.0 >

default tracks hide all add custom tracks configure refresh

.....

click → **Variation and Repeats**

SNPs (126) dense ▾	SNP Arrays hide ▾	HapMap SNPs hide ▾	HapMap LD Unph. hide ▾	Structural Var hide ▾
Segmental Dups hide ▾	Exapted Repeats hide ▾	RepeatMasker dense ▾	Interrupted Rpts hide ▾	Simple Repeats hide ▾
Microsatellite hide ▾	Self Chain hide ▾			



Get DNA in Window

Get DNA for

Position

Note: if you would prefer to get DNA for features of a particular track or

Sequence Retrieval Region Options:

Add extra bases upstream (5') and extra downstream (3')

Note: if a feature is close to the beginning or end of a chromosome and u

Sequence Formatting Options:

All upper case.
 All lower case.
 Mask repeats: to lower case to N
 Reverse complement (get '-' strand sequence)

get DNA extended case/color options

Note: The "Mask repeats" option applies only to "get DNA", not to "exte

GAAATGFAACAGTGGCTTTTAATCAATAGTACCACTCA?ACATPATGATATTCCTTAT
 ACCATGCCCTACTGCAGAAATAGTGCTACTAACCAACAGAAAAGCGTAAGATAACT
 TTAAAAACAGAAATTTGTAATGFAAACCTATATCATGCAATTTTAAATAGTAAAT
 TATCTGTGAAATTAACAGATAATACATATTTCTGTGCAATTAACACAAATATCC
 GCTTGAACCCAGGAACAGGSAATGTACATTCTTTCTTCATTTTTTGCTCCA
 TCCATCAGATAATTACTTGAAGACTgaataaatgaaaaaaaaagaccgttatac
 acaagtgaatgatagcgccttannaatagcatgcaagtaccctctggagagagag
 acttatatcatggaattatititgagcccccttttctaaanaatgtaagagagta
 aattactaaagaaaataataaaaaataccactgtaataaatttttttttttanaaatt
 tcaaatcaattataactgatactgtctattataaattctgaaaataataattcggaa
 aatagctccaattttttattataatagcttatatgacgtcactacttaataata
 tttattataatagctttttctctctctcccccacccccccatcttaaaata
 tttgaaactcactcatgctctttgagggacatggatgaagctgaaaccatcattctca

Extended DNA Case/Color

Extended DNA Case/Color Options

Use this page to highlight features in genomic DNA text. DNA covered by a particu

Position Reverse complement

Letters per line Default case: Upper Lower

Track Name	Toggle	Under-Case	line	Bold	Italic	Red	Green	Blue
Chromosome Band (Ideogram)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
Gencode Introns Oct05	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
UCSC Genes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
RefSeq Genes	<input checked="" type="checkbox"/>	<input 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"/>	<input type="checkbox"/>	0	0	0
Human mRNAs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
Spliced ESTs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
UVa DNA Rep TR50	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
TFBS Conserved	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
LI TAF1 Signal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0
SNPs (126)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	255	0
Conservation	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>	0	0	0

Red: RefSeq exon
 Green: SNP
 Yellow: exon + SNP

UCSC

Find a chicken homolog of a human protein

NCBI Entrez Protein

Search Protein for np_001455

Display: Summary Show 20 Sort by Relevance Send to

1: NP_001455.3 Reports ADAM metalloproteinase domain 2 protein [Homo sapiens]

```
>gi|55743080|ref|NP_001455.3| ADAM metalloproteinase domain 2 protein [Homo sapiens]
MWRVLFLLSGLGGLRMDNSFSLPVQITVPEKIRSIKGGISBQASYKIVIGKPYTVNLMQKNFLPHNF
RVYSYSGTQIMKPLDQDFONFCHYQYIIOYPKSVVMVSPCTGLRGLQFENYVGIPLSSVGFEBVT
YQVKKRADVLSLWEDIESRDLSEFLQVPEPODFAKYIEMHVIVEKGLYNHMGSDTTVVAQKVPFLIG
L7NAIFVSPNITILLSSLELNDENKIAATGEANELLHFTLRWKTSLVLRPHDVAFLVYREKSNVYGA
TFQKMCADANYAGGVVLPRTISLSLAVLLAQLLSLSMGIYDDINRCCSGAVCMNPEALHPSGVKIL
FNSGSPDPAHFIEKQKQCLRNQDLPFPKQAVVCNALLRAGRCOCQTEQDQALIGSTCDIATCR
FRAGSCADGPPCCNCLFMSKERMRPSPFBCDLPYCYNGSSASCPENHYVOTGHPGCLNQCIDGVCM
SGDKQCTDFGKREYFPGSBCYSHLNSKTDVSGNCGISDSGYTQCADNLQCGKLCIKYVGFLLQIPRA
TIIYANISGHLIAVEFASDHADSKHWIKDQTSQCNKVCNRQCVSSSYLVDCDCTDKNDRQVCNKK
KHCSCASYLPPDCVSDLPDGGSIDSGNFPVPAIPARLPERRYENIYHSPKMRWPFLLPFFIIFC
VLIALMVKVQKRRWRTDSSDQPESESEPKG
```

UCSC BLAT search

Home Genomes Tables PCR Session FAQ Help

Chicken BLAT Search

BLAT Search Genome

Genome: Chicken Assembly: Feb. 2004 Query type: BLAT's guess Sort output: query,score Output type: hyperlink

```
>gi|55743080|ref|NP_001455.3| ADAM metalloproteinase domain 2 protein [Homo sapiens]
MWRVLFLLSGLGGLRMDNSFSLPVQITVPEKIRSIKGGISBQASYKIVIGKPYTVNLMQKNFLPHNF
RVYSYSGTQIMKPLDQDFONFCHYQYIIOYPKSVVMVSPCTGLRGLQFENYVGIPLSSVGFEBVT
YQVKKRADVLSLWEDIESRDLSEFLQVPEPODFAKYIEMHVIVEKGLYNHMGSDTTVVAQKVPFLIG
L7NAIFVSPNITILLSSLELNDENKIAATGEANELLHFTLRWKTSLVLRPHDVAFLVYREKSNVYGA
TFQKMCADANYAGGVVLPRTISLSLAVLLAQLLSLSMGIYDDINRCCSGAVCMNPEALHPSGVKIL
FNSGSPDPAHFIEKQKQCLRNQDLPFPKQAVVCNALLRAGRCOCQTEQDQALIGSTCDIATCR
FRAGSCADGPPCCNCLFMSKERMRPSPFBCDLPYCYNGSSASCPENHYVOTGHPGCLNQCIDGVCM
SGDKQCTDFGKREYFPGSBCYSHLNSKTDVSGNCGISDSGYTQCADNLQCGKLCIKYVGFLLQIPRA
TIIYANISGHLIAVEFASDHADSKHWIKDQTSQCNKVCNRQCVSSSYLVDCDCTDKNDRQVCNKK
KHCSCASYLPPDCVSDLPDGGSIDSGNFPVPAIPARLPERRYENIYHSPKMRWPFLLPFFIIFC
VLIALMVKVQKRRWRTDSSDQPESESEPKG
```

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 25 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 PCR Session 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	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 size 186 bp.

Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

default tracks hide all add custom tracks configure refresh

Chicken BLAT Results UCSC BLAT search

Home Genomes Tables PCR Session FAQ Help

Chicken BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	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	12	385	390	735	83.4%	1	++	67659961	67659978	18

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 base mark the boundaries of gaps in either sequence.

NP_001455.3

```

mrvvlflilag lgglrmdenf dslpvqitvp ekirsiikeg iesqaasykiv iegkpytvn1 60
mqknflphnf zvyaysgtgi mkpldgdqgn ichygygyieg ypkavmvtst cttglrvqif 120
envsvyiepl eavvgefvh1 yvkhkxkdv slynkdiies rdlsikigev epegdaky1 180
emhvivekgl ynhmgsttv vaqkvfqlg ltnaifvsn itiiisletl widenkiatt 240
geanellhtf lrwktstvl rphdvaflv yreksnyvga tfggkmdan yaggvvlhpr 300
tislslavi laqlisemg ityddinkeg agavclimp eahifagvki fncsfedfa 360
hfiakqkqec lhqprldpf fkgavvagna kleageecdc gtagcdalg etcodiater 420
fkagencaeg pccencflms kerncrpafe ecdlpeycng ssaspenhy vqtghpqc1n 480
qwicldgvm agdkctdtf gkevefpgse cyhlnsktd vsngcigsda gytgccead1L 540
qORLtkiv qfllqpra TITAnisg LLavefsad hadsgkmw1 DORctienrv 600
crnqrcvss ylgdcttdk cndrgvnnk khchcsasy1 ppcavgsdl wpggsidsn 660
fppvaiarl perryieniy hskpmrvpff lfipffiiic vliamvknv fgrkkwrtd 720
ysadegpese sepgk
    
```

Chicken.chrUn :

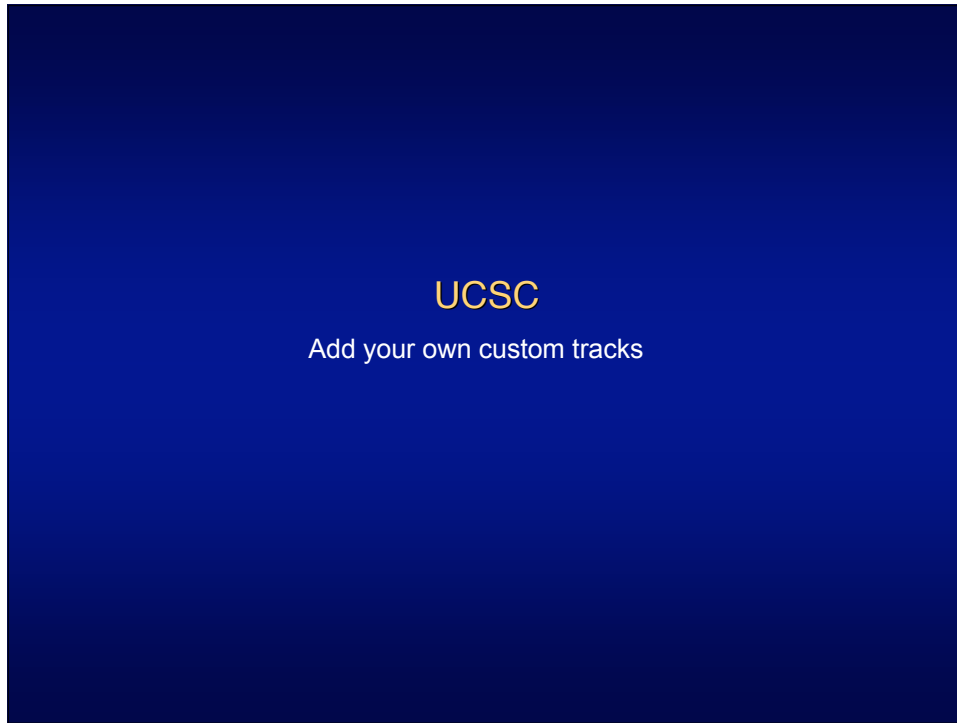
```

AATCTGGggt CTGGAAACT CATCTGCaca TA'ccaaaac gagtccctt caccaaatta 635429
aagggtACCA TCATCTATGC Tcaagtcaa gaaCATCTGT G'gtgtctt t'gatgtaatg 635489
catgcaccct cggggacaga tctctctctg gttAAGGATG GCACGaaaTG CCGT'cccgga 635549
ANGGTA
    
```

Side by Side Alignment*

```

001615 N L Q C G K L I C K Y 001647
>>>>> N G G G G G A A A A A C T C T G C A C A T A C >>>>>
635370 aatctgggctgtggaanaactcatctgcacatac 635402
    
```



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: chrX:151,073,054-151,383,976 | image width: 620 | submit

[Click here to reset](#) the browser user interface settings to their defaults.

[add custom tracks](#) | [configure tracks and display](#) | [clear position](#)

Add Custom Tracks

clade: Vertebrate | genome: Human | assembly: Mar. 2006 | [hg18]

Display your own data as custom annotation tracks in the browser. Data must be formatted in [BED](#), [GFF](#), or [BigWig](#) format. To view and display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#). Publicly available custom tracks are listed in the [User's Guide](#).

Paste URLs or data: Or upload: Browse... Submit

```

track name="Microsatellites" description="Microsatellites" color=0,128,0
chr22 38627059 38627060 D22S276
chr22 39005417 39005418 D22S307
track name="Genotyped SNPs" description="Genotyped SNPs" color=0,0,255
chr22 38518342 38518343 ss146131
chr22 38705963 38705964 ss2941443
chr22 38884157 38884158 ss141110
chr22 39171390 39171391 ss22916
    
```

Optional track documentation: Or upload: Browse... Clear

Click [here](#) for an HTML document template that

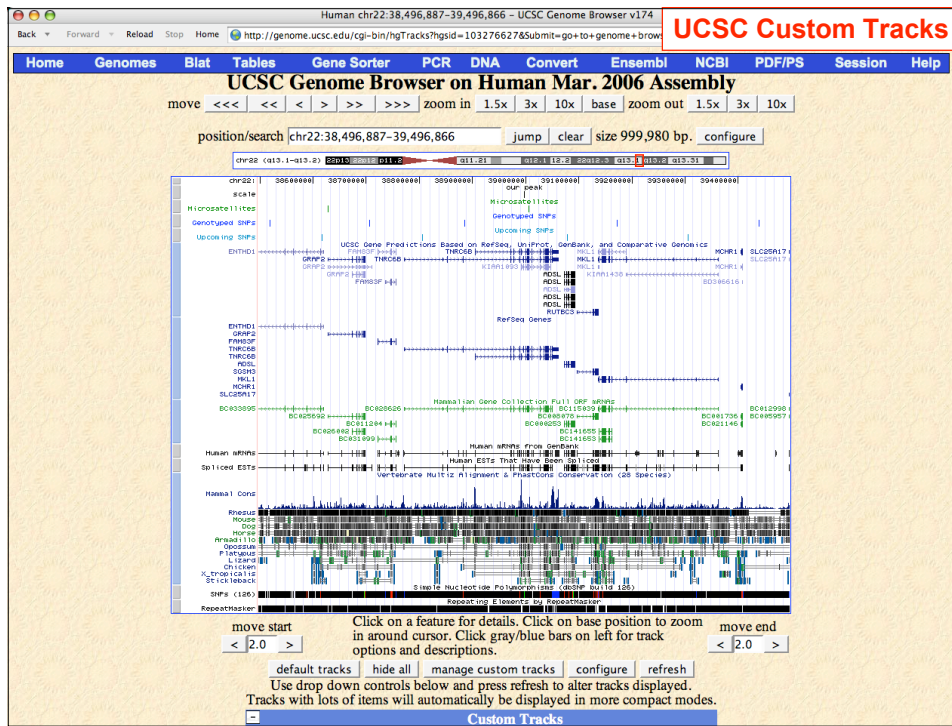
Manage Custom Tracks

genome: Human | assembly: Mar. 2006 | [hg18]

Name	Description	Type	Doc	Items	Pos	delete	
Upcoming SNPs	Upcoming SNPs	bed		4	chr22:	<input type="checkbox"/>	add custom tracks
Genotyped SNPs	Genotyped SNPs	bed		5	chr22:	<input type="checkbox"/>	go to genome browser
Microsatellites	Microsatellites	bed		2	chr22:	<input type="checkbox"/>	go to table browser
scale	our peak	bed		1	chr22:	<input type="checkbox"/>	

check all / clear all

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 retrieve the data associated with a track in text format, to calculate intersections between track track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and want to use [Galaxy](#) or our [public MySQL server](#). Refer to the [Credits](#) page for the list of contributors and usage r

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 define regions

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

correlation: create

output format: all fields from selected table Send output to Galaxy

output file: (leave blank to keep output in browser)

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 *

chrom does match *

strand does match *

txStart is ignored

txEnd is ignored

cdsStart is ignored

cdsEnd is ignored

exonCount is = 1

exonStarts does match *

exonEnds does match *

id is ignored

name2 does match *

cdsStartStat does match *

cdsEndStat does match *

exonFrames does match *

#	bin	name	chrom	strand	txStart	txEnd	cdsStart	cdsEnd	exonCount	exonStarts	exonEnds	id	name2	cdsStartStat	cdsEndStat	exonFrames
585	NM_0010054	chr1	+	58953	59871	58953	59871	1	58953	59871	0	OR4F5	cmpl	cmpl	0	
587	NM_0010052	chr1	+	357521	358458	357521	358458	1	357521	358458	0	OR4F3	cmpl	incmpl	0	
587	NM_0010052	chr1	+	357521	358458	357521	358458	1	357521	358458	0	OR4F6	cmpl	incmpl	0	
587	NM_0010052	chr1	+	357521	358460	357521	358460	1	357521	358460	0	OR4F29	cmpl	cmpl	0	
589	NM_0010052	chr1	-	610958	611897	610958	611897	1	610958	611897	0	OR4F29	cmpl	cmpl	0	
589	NM_0010052	chr1	-	610960	611897	610960	611897	1	610960	611897	0	OR4F3	incmpl	cmpl	0	
589	NM_0010052	chr1	-	610960	611897	610960	611897	1	610960	611897	0	OR4F16	incmpl	cmpl	0	
593	NM_080005	chr1	+	1157491	1160281	1157521	1158511	1	1157491	1160281	0	BIGALF6	cmpl	cmpl	0	
607	NM_080431	chr1	+	2927905	2929325	2928110	2929244	1	2927905	2929325	0	ACTR2	cmpl	cmpl	0	
88	NM_006511	chr1	+	15858950	15860803	15858950	15860803	1	15858950	15860803	0	KSC1A1	cmpl	incmpl	0	
707	NM_0010895	chr1	-	16006243	16006781	16006455	16006731	1	16006243	16006781	0	LOC440567	cmpl	cmpl	0	

NCBI

View a genomic region between two STS markers

NCBI Home Page

National Center for Biotechnology Information
National Library of Medicine National Institutes of Health

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search All Databases 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

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

GenBank Celebrating 25 Years
NCBI will hold a scientific meeting to celebrate the 25th anniversary of GenBank. April 7-8, 2008. Natcher Auditorium, NIH Campus, Bethesda MD. [click here for more information](#)

GenBank vs. RefSeq
Confused about the distinctions between GenBank, RefSeq, TRS and UniProt? [Click here for a brief description of the databases and their differences.](#)

PubMed Central
An archive of biomedical and life sciences journals.
Free fulltext
Over 1,100,000 articles from over 340 journals
Linked 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

NCBI Map Viewer

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

Search
Search: Homo sapiens
for: d8s1170 OR d8s94
Go

Tools Legend
Search or Browse the Genome
BLAST
Genome Resources page

News
Annotation update plans for Human Genome build 36 Oct 1, 2007
NCBI is planning to update its annotation of the human genom... [more](#)
[Show all](#)

Related Resources
NCBI Home
NCBI Web Search
NCBI Site map
Genome Biology
Taxonomy
Entrez (Global Query)
BLAST
Map Viewer FTP

Small Genomes
Bacteria
Organelles
Viruses

Scientific name	Common name	Build	Tools
Vertebrates (15)			
Mammals (13)			
Primates (3)			
<i>Homo sapiens</i>	human	Build 36.2 Build 35.1	ⓐ ⓑ ⓓ
<i>Macaca mulatta</i>	rhesus macaque	Build 1.1	ⓐ ⓑ ⓓ
<i>Pan troglodytes</i>	chimpanzee	Build 2.1	ⓐ ⓑ ⓓ
Rodents (2)			
<i>Mus musculus</i>	laboratory mouse	Build 37.1 Build 36.1	ⓐ ⓑ ⓓ
<i>Rattus norvegicus</i>	rat	RGSC v3.4	ⓐ ⓑ ⓓ
Monotremes (1)			
Marsupials (1)			
Other Mammals (6)			
Other Vertebrates (2)			
Invertebrates (7)			
Protozoa (7)			
Plants (42)			
Fungi (16)			
<i>Aspergillus fumigatus</i>		Build 2.1	ⓐ ⓑ ⓓ
<i>Aspergillus niger</i>		Build 1.1	ⓐ ⓑ ⓓ
<i>Candida glabrata</i>		Build 1.1	ⓐ ⓑ
<i>Cryptococcus neoformans</i>		Build 2.1	ⓐ ⓑ
<i>Debaryomyces hansenii</i>		Build 1.1	ⓐ ⓑ
<i>Encephalitozoon cuniculi</i>		Build 1.1	ⓐ ⓑ
<i>Eremothecium gossypii</i>		Build 3.1	ⓐ ⓑ
<i>Gibberella zeae</i>		Build 1.2	ⓐ ⓑ
<i>Kluyveromyces lactis</i>		Build 1.1	ⓐ ⓑ
<i>Magnaporthe oryzae</i>	rice blast fungus	Build 3.1	ⓐ ⓑ
<i>Neurospora crassa</i>		Build 1.1	ⓐ ⓑ
<i>Pichia stipitis</i>		Build 1.1	ⓐ ⓑ
<i>Saccharomyces cerevisiae</i>	baker's yeast	Build 2.1	ⓐ ⓑ ⓓ
<i>Schizosaccharomyces pombe</i>	fission yeast	Build 1.1	ⓐ ⓑ
<i>Ustilago maydis</i>		Build 1.1	ⓐ ⓑ
<i>Yarrowia lipolytica</i>		Build 1.1	ⓐ ⓑ

Entrez Genome view
http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606&query=d8s1170&20OR&2

NCBI NCBI Map Viewer

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

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

BLAST search the human genome

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

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

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

NCBI NCBI Map Viewer

Search: Find Find in This View Advanced Search

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 [clear]

Master Map: STS Summary of Maps Maps & Options
Region Displayed: 32M-44,400K bp Download/View Sequence/Evidence

Marker	STS	Gene	Map	Go to
RH120054	●			Info
STS-N33323	●			Info
G29276	●			Info
D8S1170	●			Info
RH36053	●			Info
SHGC-110285	●			Info
SHGC-156087	●			Info
RH122187	●			Info
AFMB283XF1	●			Info
D8S94	●			Info
RH67241	●			Info

Summary of Maps:

- Map 1: Homo sapiens UniGene Clusters
Region Displayed: 32M-44,400K bp
Total UniGene Clusters On Chromosome: 28983 [149 not localized]
UniGene Clusters Labeled: 50 Total UniGene Clusters in Region: 1656
- Map 2: Genes On Sequence
Region Displayed: 32M-44,400K bp
Total Genes On Chromosome: 984 [6 not localized]
Genes Labeled: 50 Total Genes in Region: 85
- Map 3: STS
Region Displayed: 32M-44,400K bp
Total STSs On Chromosome: 8517 [11 not localized]
STSs Labeled: 30 Total STSs in Region: 470



NCBI Maps & Options

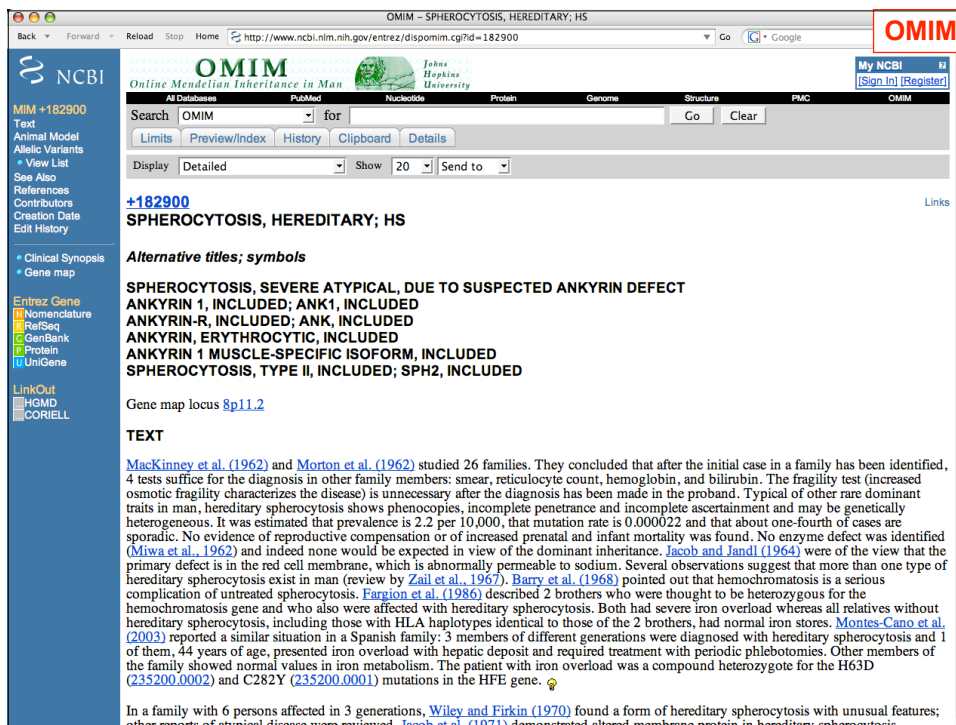
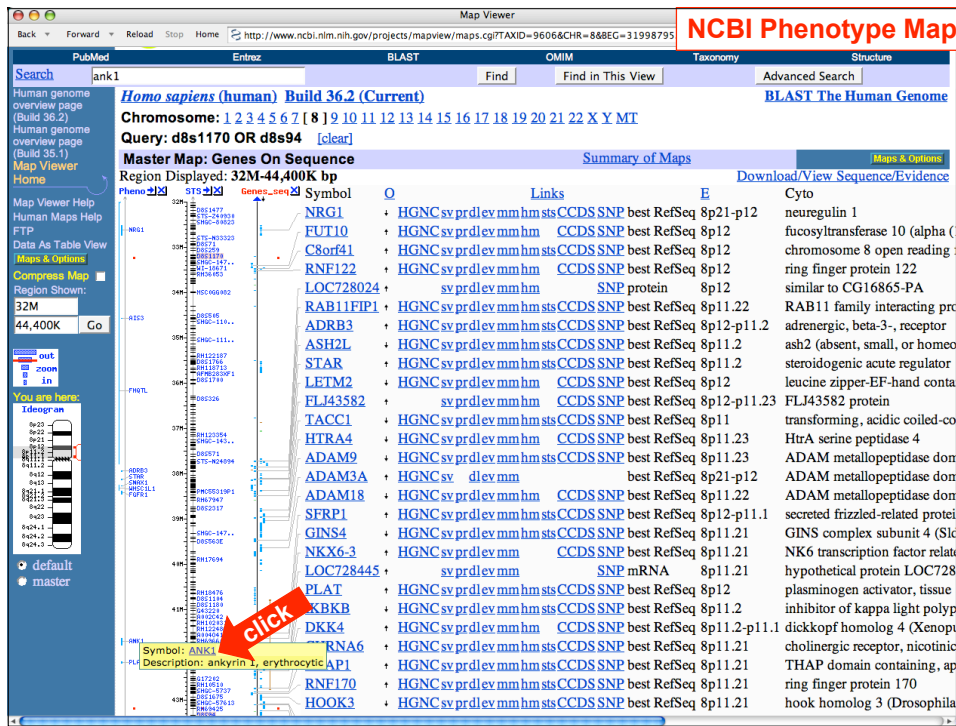
click

Organism: Homo sapiens
Region Shown: 31998795.3|44415851.6

Available Maps:
Org: human Assembly: reference
Ensembl Transcripts
GenBank DNA
Gene
NCI Clone
Phenotype
RefSeq Transcripts
Repeats
rnaBt
rmaGsa

Maps Displayed (left to right):
[] Phenotype
[R] STS
[] Gene

More Options:
 Show Connections Verbose Mode
Compress Map: off Auto Compress if > 350 px
Page Length: 30
Thumbnail View: default (ideogram) master



Map Viewer
http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?TAXID=9606&C

NCBI region between 2 genes

Human genome overview page (Build 35.1)
Map Viewer Home

Human genome overview page (Build 35.2)
Human genome overview page (Build 35.1)
Map Viewer Home

Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map

Region Show: Go

You are here: Go

Zoom: out in

default master

Homo sapiens (human) 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 [clear]

Master Map: Genes On Sequence
Region Displayed: 32M-44,400K bp

Summary of Maps
Download/View Sequence/Evidence

Pheno	STS	Gene_seq	Symbol	Q	Links	E	Cyto
			NRG1	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p21-p12	neuregulin 1
			FUT10	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p12	fucosyltransferase 10 (alpha (
			C8orf41	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p12	chromosome 8 open reading
			RNF122	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p12	ring finger protein 122
			LOC728024	+	sv pr d lev mm hm sts	SNP protein 8p12	similar to CG16865-PA
			RAB11FIP1	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.22	RAB11 family interacting pro
			ADRB3	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p12-p11.2	adrenergic, beta-3-, receptor
			ASH2L	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.2	ash2 (absent, small, or homeo
			STAR	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.2	steroidogenic acute regulator
			LETM2	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p12	leucine zipper-EF-hand conta
			FLJ43582	+	sv pr d lev mm hm sts	CCDS SNP best RefSeq 8p12-p11.23	FLJ43582 protein
			TACC1	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11	transforming, acidic coiled-co
			HTRA4	+	HGNC sv pr d lev mm hm sts	CCDS SNP best RefSeq 8p11.23	Htra serine peptidase 4
			ADAM9	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.23	ADAM metallopeptidase dom
			ADAM3A	+	HGNC sv d lev mm	best RefSeq 8p21-p12	ADAM metallopeptidase dom
			ADAM18	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.22	ADAM metallopeptidase dom
			SFRP1	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p12-p11.1	secreted frizzled-related protei
			GINS4	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.21	GINS complex subunit 4 (Sik
			NKX6-3	+	HGNC sv pr d lev mm	CCDS SNP best RefSeq 8p11.21	NK6 transcription factor relat
			LOC728445	+	sv pr d lev mm	SNP mRNA 8p11.21	hypothetical protein LOC728
			PLAT	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p12	plasmidogen activator, tissue
			IKBKB	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.2	inhibitor of kappa light poly
			DKK4	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.2-p11.1	dickkopf homolog 4 (Xenopi
			CHRNA6	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.21	cholinergic receptor, nicotinic
			THAP1	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.21	THAP domain containing, ap
			RNF170	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.21	ring finger protein 170
			HOOK3	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.21	hook homolog 3 (Drosophila
			FLJ23356	+	sv pr d lev mm hm	CCDS SNP best RefSeq 8p11.21	hypothetical protein FLJ2335
			LOC41347	+	sv pr d lev mm hm	SNP protein 8p11.21	similar to family with sequenc

click

Map Viewer
http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?TAXID=9606&C

NCBI region between 2 genes

Human genome overview page (Build 35.1)
Map Viewer Home

Human genome overview page (Build 35.2)
Human genome overview page (Build 35.1)
Map Viewer Home

Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map

Region Show: Go

You are here: Go

Zoom: out in

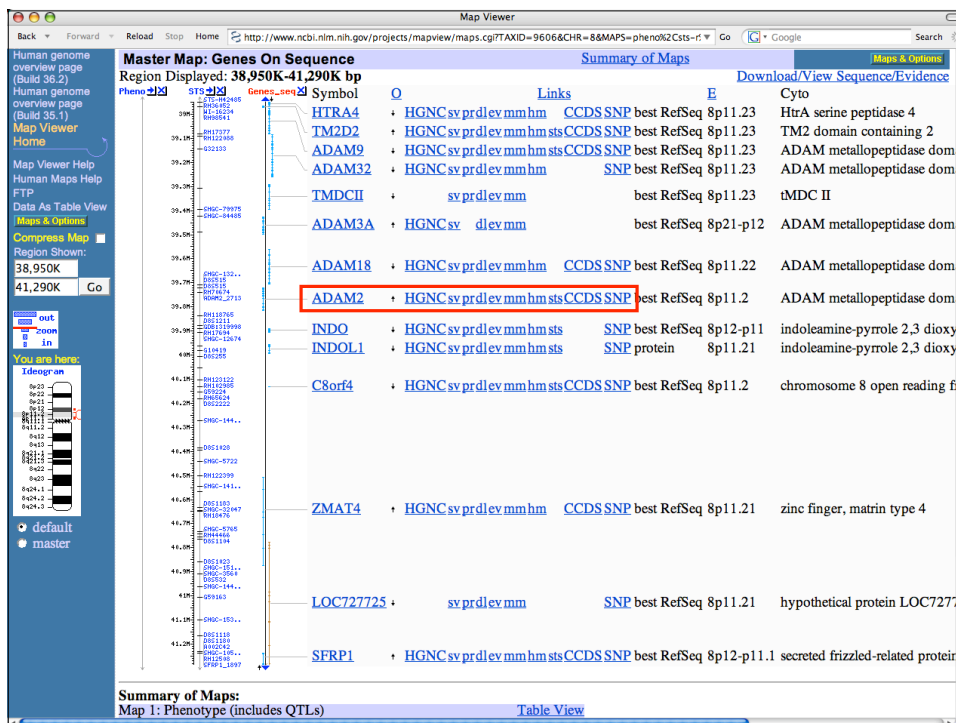
default master

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

Summary of Maps
Download/View Sequence/Evidence

Pheno	STS	Gene_seq	Symbol	Q	Links	E	Cyto
			HTRA4	+	HGNC sv pr d lev mm hm	CCDS SNP best RefSeq 8p11.23	Htra serine peptidase 4
			TM2D2	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.23	TM2 domain containing 2
			ADAM9	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.23	ADAM metallopeptidase dom
			ADAM32	+	HGNC sv pr d lev mm hm	SNP best RefSeq 8p11.23	ADAM metallopeptidase dom
			TMDCC1	+	sv pr d lev mm	best RefSeq 8p11.23	tMDC II
			ADAM3A	+	HGNC sv d lev mm	best RefSeq 8p21-p12	ADAM metallopeptidase dom
			ADAM18	+	HGNC sv pr d lev mm hm	CCDS SNP best RefSeq 8p11.22	ADAM metallopeptidase dom
			ADAM2	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.2	ADAM metallopeptidase dom
			INDO	+	HGNC sv pr d lev mm hm sts	SNP best RefSeq 8p12-p11	indoleamine-pyrole 2,3 dioxy
			INDOL1	+	HGNC sv pr d lev mm hm sts	SNP protein 8p11.21	indoleamine-pyrole 2,3 dioxy
			C8orf4	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p11.2	chromosome 8 open reading f
			ZMAT4	+	HGNC sv pr d lev mm hm	CCDS SNP best RefSeq 8p11.21	zinc finger, matrin type 4
			LOC727725	+	sv pr d lev mm	SNP best RefSeq 8p11.21	hypothetical protein LOC727
			SFRP1	+	HGNC sv pr d lev mm hm sts CCDS SNP	best RefSeq 8p12-p11.1	secreted frizzled-related protei

Summary of Maps:
Map 1: Phenotype (includes QTLs) Table View



Entrez Gene: ADAM2 ADAM metallopeptidase domain 2 (fertilin beta) [Homo sapiens]

Search Gene for [] Go Clear

Display Full Report Show 20 Send to

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

1: ADAM2 ADAM metallopeptidase domain 2 (fertilin beta) [Homo sapiens] updated 17-Jan-2008

GeneID: 2515

Summary

Official Symbol ADAM2 provided by HGNC

Official Full Name ADAM metallopeptidase domain 2 (fertilin beta) provided by HGNC

Primary source HGNC:198

See related Ensembl:ENSG00000104755; HPRD:03322; MIM:601533

Gene type protein coding

RefSeq status Reviewed

Organism [Homo sapiens](#)

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo

Also known as 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) Go to [reference sequence details](#) [Try our new Sequence Viewer](#)

NC_000008.9

Entrez Gene: ADAM2 ADAM metallopeptidase domain 2 (fertilin beta) [Homo sapiens]

NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

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

mRNA and Protein(s)

1. **NM_001464.3--NP_001455.3 ADAM metallopeptidase domain 2 proprotein**

Source sequence(s) BG719616.B1460477.U38805

Consensus CDS CCDS34884.1

Consensus CDS O29965

UniProtKB/Swiss-Prot

Conserved Domains (5) **summary**

smart00650	DISIN; Homologues of snake disintegrins; Snake disintegrins inhibit the binding of ligands to integrin receptors. They contain a 'RGD' sequence, identical to the recognition site of many adhesion proteins.
smart00608	ACR; ADAM Cysteine-Rich Domain;
cd04269	ZnMc_adamalsyn_II_like; Zinc-dependent metalloprotease; adamalsyn_II_like subfamily. Adamalsyn II is a snake venom zinc endopeptidase. This subfamily contains other snake venom metalloproteases, as well as membrane-anchored metalloproteases belonging to the ADAM family.
pfam00200	Disintegrin; Disintegrin.
pfam01562	Pop_M12B_propep; Reprolysin family propeptide. This region is the propeptide for members of peptidase family M12B. The propeptide contains a sequence motif similar to the "cysteine switch" of the matrixins.

RefSeqs of Annotated Genomes: Build 36.2

The following sections contain reference sequences that belong to a specific genome build. [Explain](#)

Reference assembly

Genomic

1. **NC_000008.9 Reference assembly**

Range 39814936..39720411, complement

Download [GenBank](#) [FASTA](#) [Sequence Viewer \(beta\)](#)

2. **NT_007995.14**

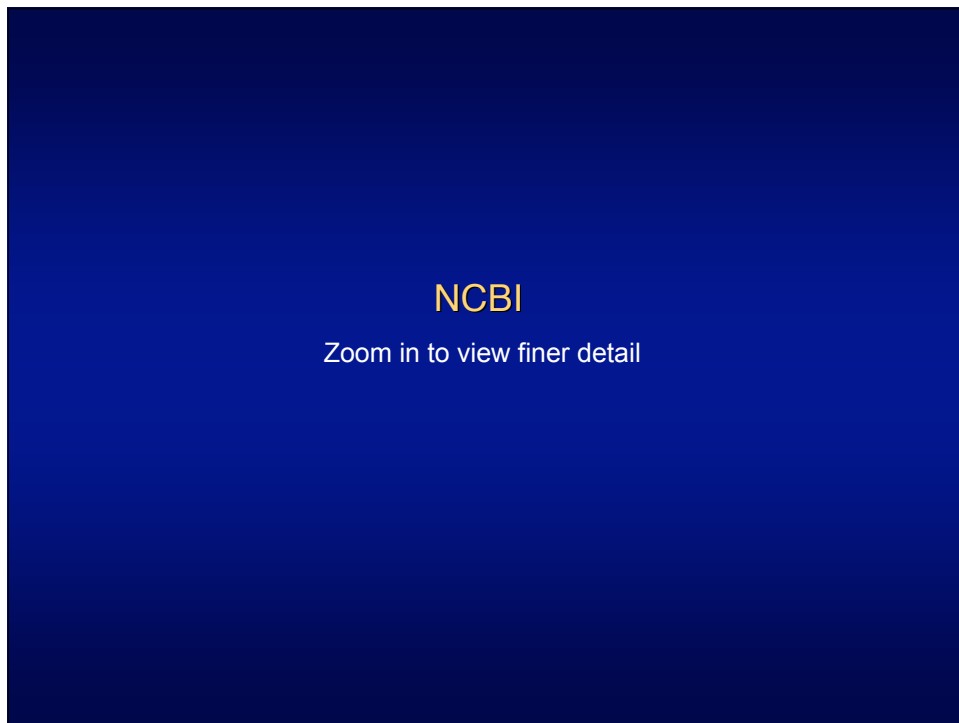
Range 10016168..9921643, complement

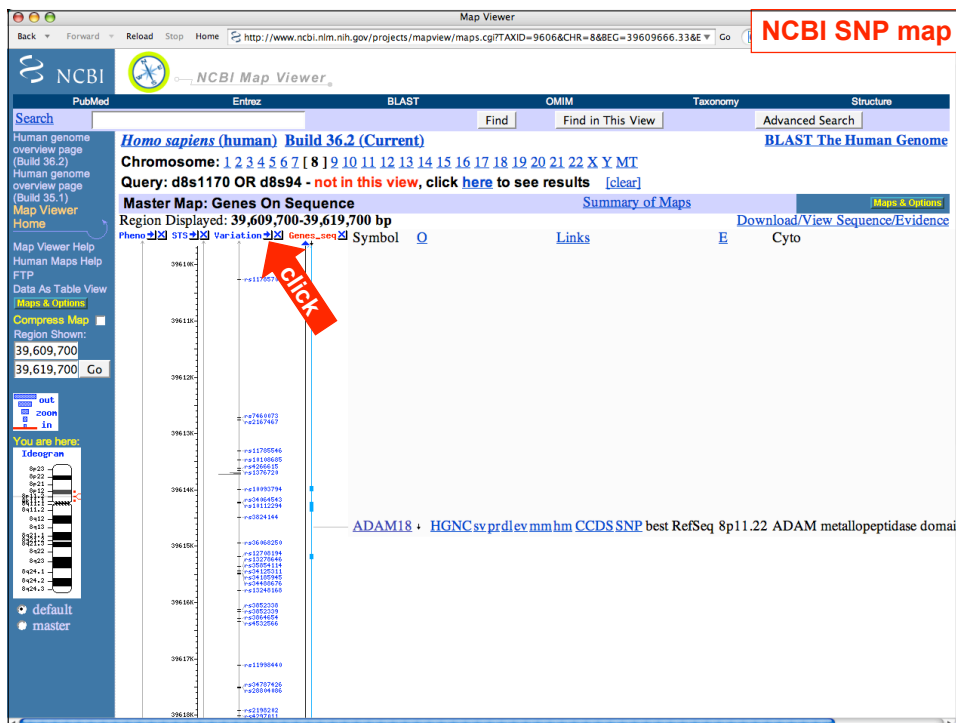
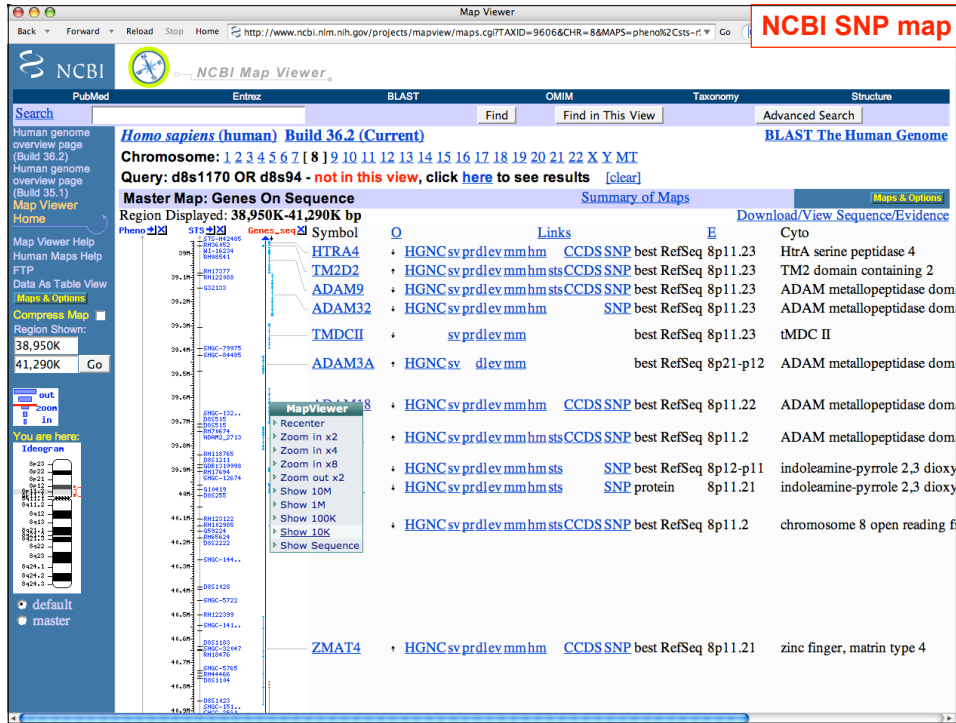
Download [GenBank](#) [FASTA](#) [Sequence Viewer \(beta\)](#)

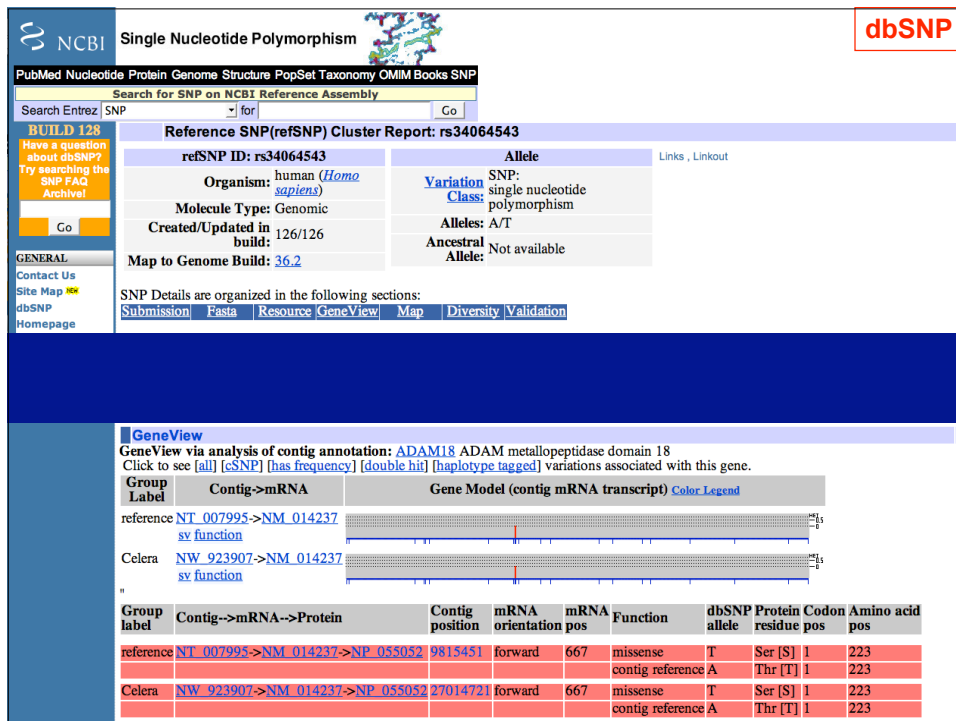
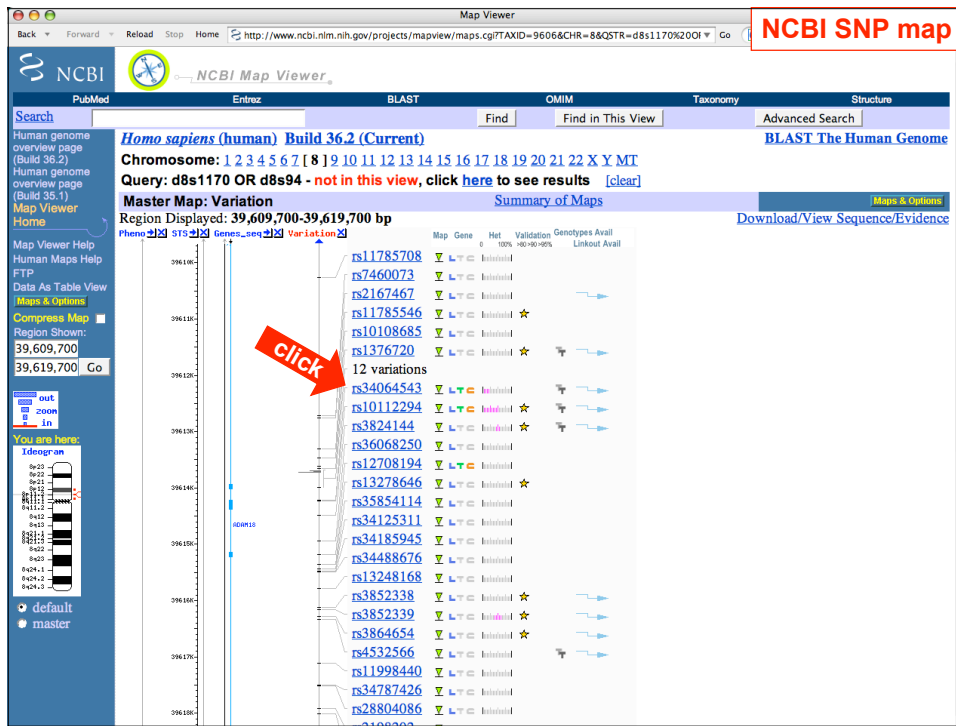
Alternate assembly (based on Celera assembly)

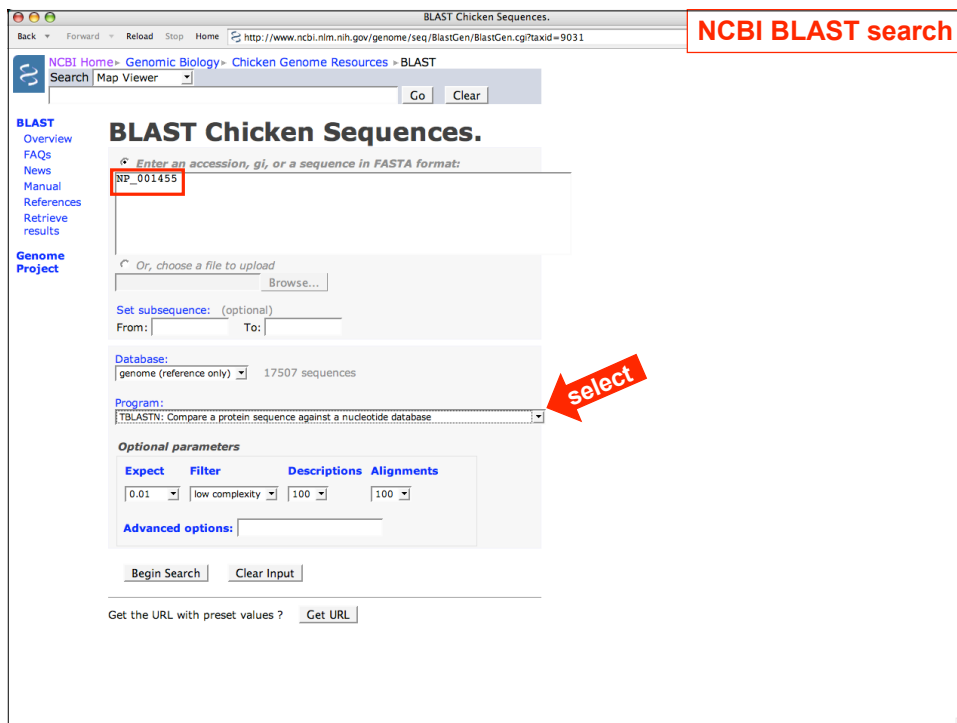
Genomic

The screenshot shows the HomoloGene web application interface. At the top, there is a search bar with 'HomoloGene' entered and a 'Go' button. Below the search bar, there are options for 'Limits', 'Preview/Index', 'History', and 'Clipboard'. The main content area displays search results for 'HomoloGene:1127. Gene conserved in Eutheria'. It lists several genes and proteins, including ADAM2 from Homo sapiens, Pan troglodytes, and Canis lupus familiaris. The interface also includes sections for 'Alignment Scores' and 'Conserved Domains'. A 'BLAST' button is visible in the alignment section.









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 metalloproteinase domain 2 propeptide (Homo sapiens) >gi|2820251|sp|Q99965|ADAM2_HUMAN ADAM 2 precursor (A disintegrin and metalloproteinase domain 2) (Fertilin subunit beta) (PH-30) (PH30) >gi|4585655|emb|CA040812.1| Fertilin beta protein [Homo sapiens] >gi|119583677|gb|EAM63273.1| ADAM metalloproteinase domain 2 (fertilin beta), isoform CRA_a [Homo sapiens]
Length=735

Distribution of 39 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

Color key for alignment scores

<40	40-50	50-80	80-200	>=200
-----	-------	-------	--------	-------

Sequences producing significant alignments:	Score (Bits)	E Value
ref NW_001471710.1 Gga5_WGA136_2 Gallus gallus chromosome 5 g...	359	1e-97
ref NW_001471459.1 Gga15_WGA207_2 Gallus gallus chromosome 15...	273	9e-72
ref NW_001471578.1 Gga22_WGA768_2 Gallus gallus chromosome 22...	63.2	4e-18

> [ref|NW_001471710.1|Gga5_WGA136_2](#) D Gallus gallus chromosome 5 genomic contig, reference assembly (based on Gallus_gallus-2.1)
Length=32733142

Query 463	ASCPENHYVQCHPCGLNQHICIDVCHSGDQCTDFGKEVFPSPCYSHLSKTDVSG	522
	C E+ YVQ G PC + C G C S KQ PG++ P BC+ +N++ D	
Sbjct 214257	PMQEDLYVDGTFPCSDGAY-CYKGCSSHSKCKHLFRQARPALECFKVNTRGDRF	214433
Query 523	GNGGIEDS-GVTQCADNLCQCKLICKYVGFLLQIPRATIIYANISGHLCAVEFASDH	581
	GNGG ++ +t+c +N CG+L C+ V + L 2+I G C +++	
Sbjct 214434	GNGSRNNIRFTKCSVENALCGRCQENVRHLLPLQNHITVIQTDPAGKNCWGLDYHLOV	214613
Query 582	ADSQKNHIDQTSCSBNKVCNRQCVSSVLYDCTCKCNRGVNCKKHCCHCASVYLP	641
	S ++DGF+CGSNK+C + C + S L YDC KC+RQVCNN R+CIC + P	
Sbjct 214614	PTSDWGAVEDGTTCSGNKICIERCTNIEVLNDCNITKCHNRGVNCKKHCCKYGNAP	214793

Gallus gallus (chicken) genome view

NCBI BLAST search

Build 2.1 statistics

Hit GIs:
Hits: 1 2 3 4 5 6 7 8 9 10 11
0 -0- 1 1 1 1 1 1 1 1 1 1
14 15 16 17 18 19 20 21 22 23 24
Hits: 1 2 2 1
27 28 29 30 31 32 33 34 35 36 37
LGE64 LGE22C19W28 E50C23 HIT not placed
Hit GIs:
Hits: 2 2

Color key for scores: <40 40-50 50-80 80-200

Gallus gallus (chicken) Build 2.1

Chromosome: 1 2 3 4 [5] 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
LGE22C19W28 E50C23 MT

Query: BLAST: ref|np_001455| (clear)

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

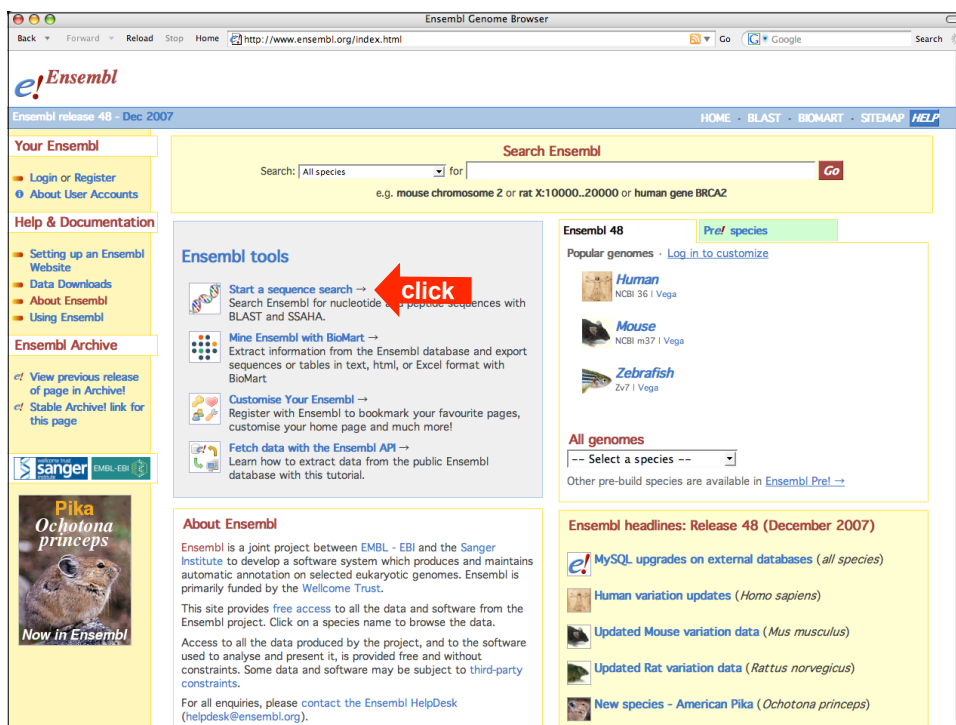
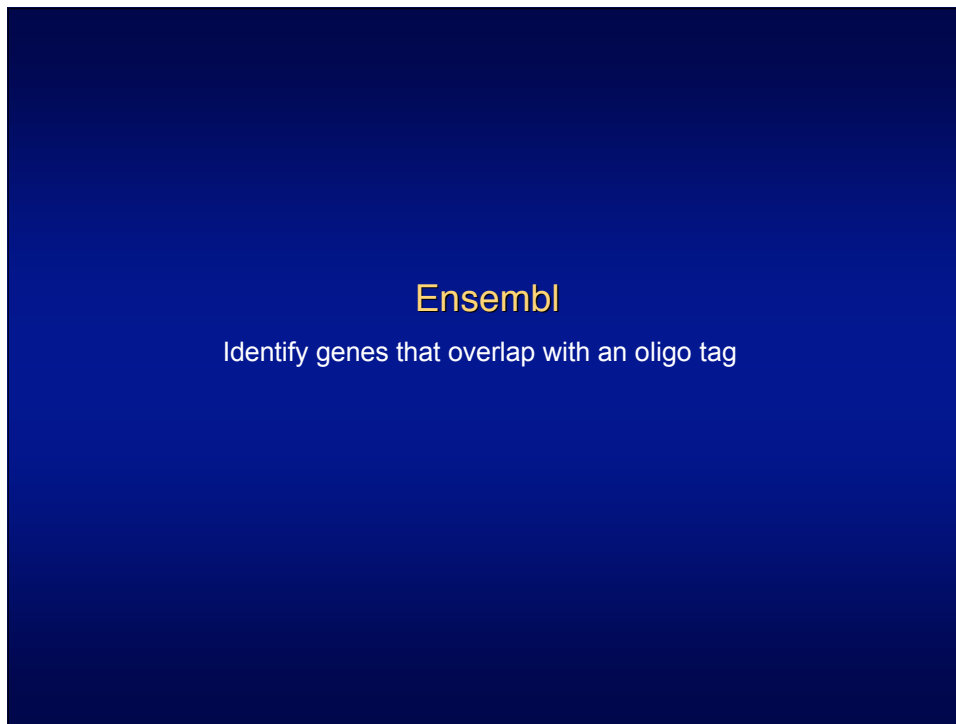
Master Map: Contig

Region Displayed: 29,718,390-29,720,890 bp

Model [x](#) RefSeq [x](#) Rfam [x](#) Genes_seq [x](#) Contig [x](#) Symbol [Q](#)

297184000
297185000
297186000
297187000
297188000
297189000
297190000
297191000
297192000
297193000
297194000
297195000
297196000
297197000
297198000
297199000
297200000
297201000
297202000
297203000

NW_001471710.1 ↓
Blast hit Identity=31% 662.4



Ensembl BLAST search

Ensembl release 48 - Dec 2007

HOME · BLAST · BIOMART · SITEMAP · HELP

Your Ensembl

Login or Register
About User Accounts

sanger EMBL-EBI

Pika Ochotona princeps
Now in Ensembl

new SETUP CONFIG RESULTS DISPLAY refresh Online Help

Enter the Query Sequence

Either Paste sequences (max 30 sequences) in FASTA or plain text:
MPSS 1
AAAAAATGTCGGCCTGAAGAG

Or Upload a file containing one or more FASTA sequences
Browse...

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

Or Enter an existing ticket ID:
Retrieve

dna queries
 peptide queries

Select the databases to search against

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

Gallus_gallus
Gasterosteus_acleatus
Homo_sapiens

dna database
 peptide database
Known Consensus CDS Peptides (CCDS peptides)

Select the Search Tool

BLASTN
BLAT
SSAHA2

Search sensitivity:
Optimise search parameters to find the following alignments

Near-exact matches (oligo)
Exact matches
Near-exact matches
Near-exact matches (oligo)
Allow some local mismatch
Distant homologies
No optimisation

Summary

setup
Not yet initialised

configure
Not yet initialised

results
Not yet initialised

display
Not yet initialised

About BlastView

BlastView provides an integrated offering access to both BLAST against Ensembl databases,

Ensembl BLAST search

Mouse Lemur Microcebus murinus
Now in Ensembl

100% identity over 100% of the query length

Query

Start	End	% ID
1	20	100.00

Alignment Locations vs. Query (click arrow to hide)

+ track coverage
->MPSS_1

Alignment Summary (click arrow to hide)

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

Subject	Chromosome	Supercontig	Clone	Contig	Chromosome	Start	End	Ori	Score	E-val	PWD	Length		
[A] [S] [G] [C]	1	20	+	Chr:15	54998168	54998187	+	Chr:15	54998168	54998187	20	0.0063	100.00	20
[A] [S] [G] [C]	1	17	+	Chr:8	72205113	72205129	-	Chr:8	72205113	72205129	17	0.39	100.00	17

click

refresh

Genomic sequence

BLASTN

Oligo

sensitivity

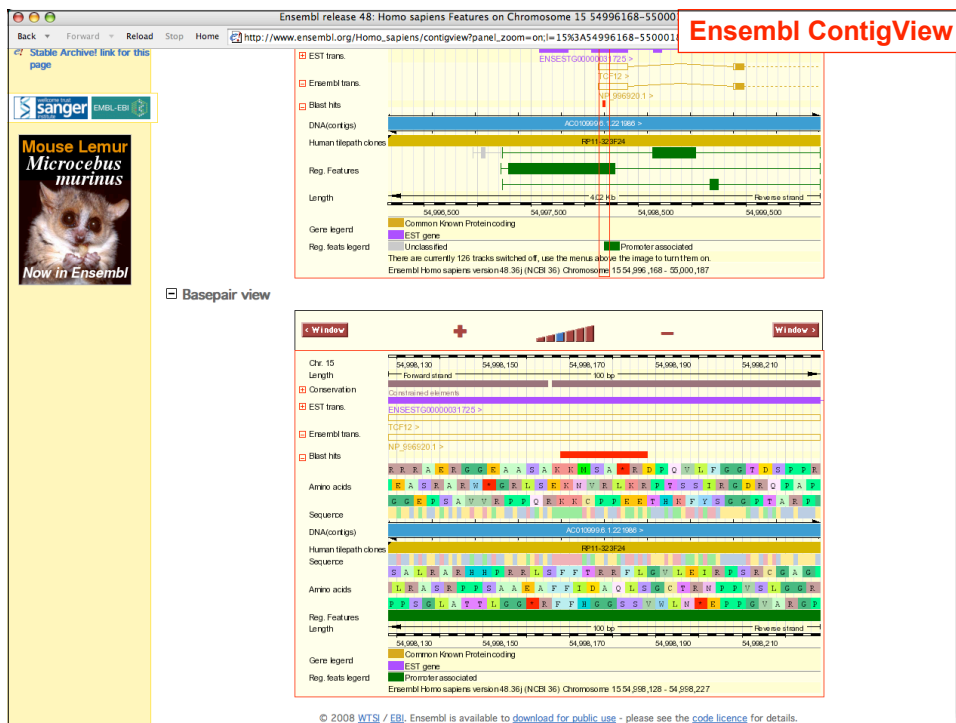
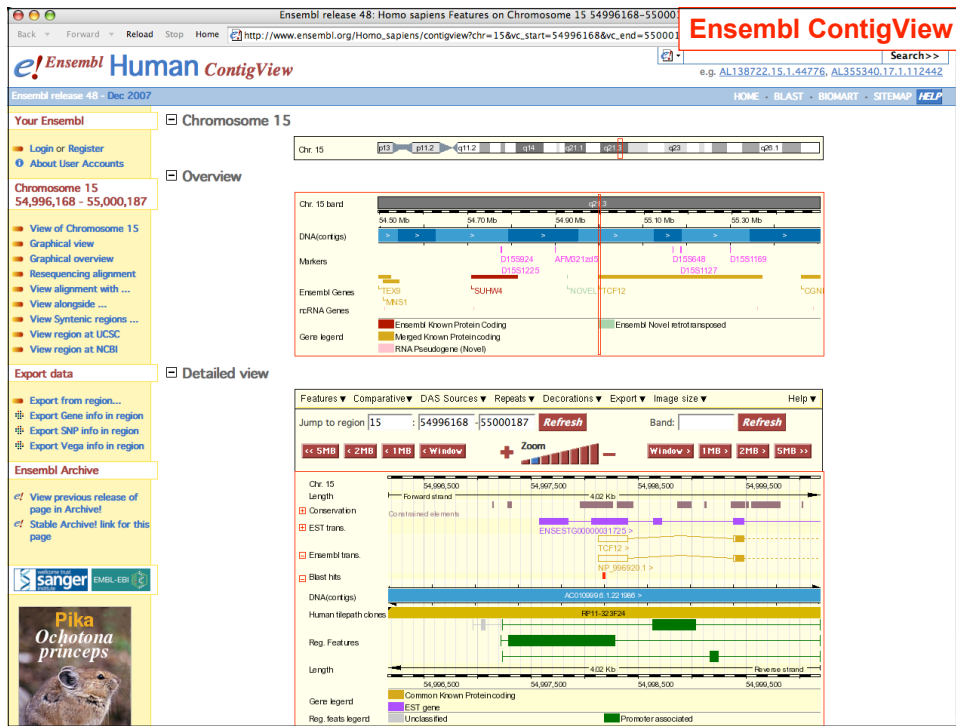
configure

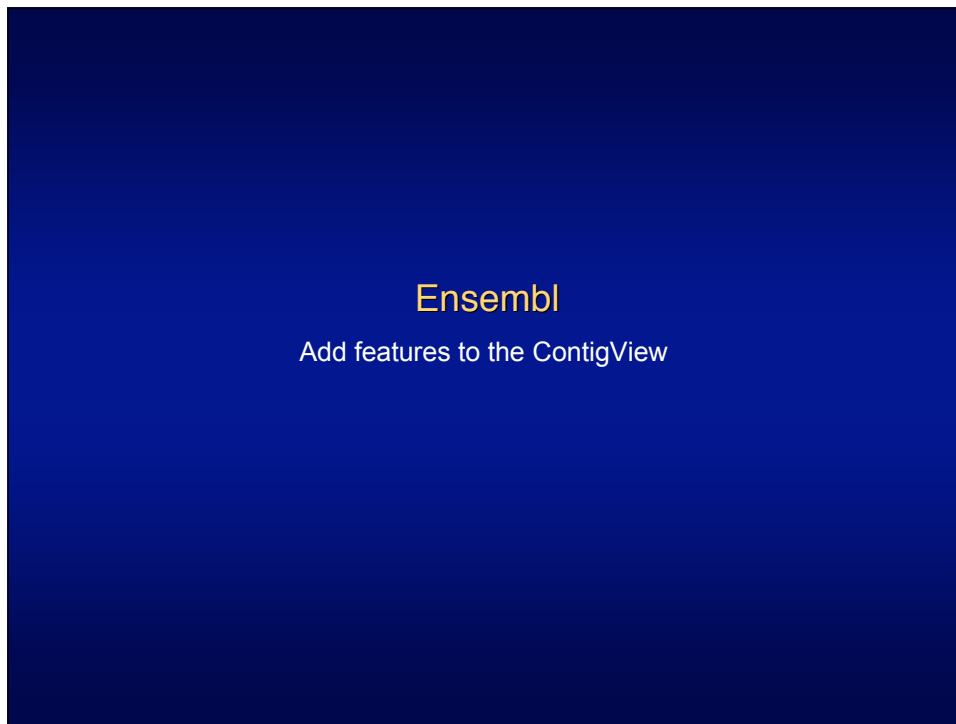
-E: 10
-B: 100
-filter: none
-RepeatMasker
-W: 11
-M: 1
-N: -3
-Q: 3
-R: 3

results

display

Not yet initialised





Ensembl release 48: Homo sapiens Features on Chromosome 15 54996168-55000187

Ensembl ContigView

Ensembl release 48 - Dec 2007

HOME · BLAST · BIOMART · SITEMAP · HELP

Your Ensembl

- Login or Register
- About User Accounts

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

- View of Chromosome 15
- Graphical overview
- Resequencing alignment
- View alignment with ...
- View alongside ...
- View Syntenic regions ...
- View region at UCSC
- View region at NCBI

Export data

- Export from region...
- Export Gene info in region
- Export SNP info in region
- Export Vega info in region

Ensembl Archive

- View previous release of page in Archive!
- Stable Archive! link for this page

Sanger EMBL EBI

Pika Ochotona princeps

Chromosome 15

Chr 15 p12 p11.2 q14 q21.1 q23 q24

Overview

Detailed view

select

Features

- SNPs
- Genotyped SNPs
- Affy 100k SNP
- Affy 500k SNP
- Regulatory features
- CTCF
- IRNA
- CpG islands
- Epigenetic regions
- FISH EF
- Markers
- isoRED/ncRNA
- isoRED search regions
- Vega assembly
- Ensembl genes
- Ig segments
- Vega Havana gene
- Vega External gene
- mRNA genes
- EST genes
- GenScan
- UniProtKB_SW
- UniProtKB_TR
- UniProtKB (non-membr)
- UniProtKB (non-membr)
- Human proteins
- Other proteins
- Unigene
- ENSEMBL mRNAs
- MiRNA
- iFam
- Human cDNAs
- cDNAs
- Sheep BAC ends
- Human ESTs
- ESTs
- Human EST
- ESTs

Forward strand

Reverse strand

Common Known Protein-coding

EST gene

Uncharacterized

are currently 126 tracks switched off, use the menus above the image to turn them on

emb1 Homo sapiens version 48.36j (NCBI 36) Chromosome: 15 54,996,168 - 55,000,187

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

Ensembl ContigView

Ensembl release 48 - Homo sapiens Features on Chromosome 15 54996168-55000187

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

Features Comparative DAS Sources Repeats Decorations Export Image size Help

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

Chr 15
Length 402 kb
Conserved elements
ENST00000400311.2 S
TCF12
V3_366520.1

DNA (contigs)
AC030989.1:221590

Human (path clones)
P11-22324

SNP: rs12915002 **click**

Reg Features
Length 402 kb
Common Known Protein coding
EST gene
Upstream
Intergenic
Intronic
Promoter associated

Gene legend
SNP legend
Reg. feats legend

There are currently 125 tracks switched off, use the menus above the image to turn them on.
Ensembl Homo sapiens version 48.36 (NCBI 36) Chromosome: 15:54,996,168 - 55,000,187

© 2008 WTS / EB. Ensembl is available for download for public use - please see the code licence for details.

Ensembl SNPView

SNPView
http://www.ensembl.org/Homo_sapiens/snpview?snp=rs12915002;source=dbSNP;c=15:54999139

54,999,139

Flanking sequence

SNP rs12915002 is located in the following transcripts

Genomic location (strand)	Gene	Transcript relative SNP position	Translation relative SNP position	Type	GeneSNPView
15: 54999139-54999139 (1)	ENSG00000140262	ENST00000333725: n/a	ENSP00000331057: n/a	INTRONIC	SNPs in gene context
	ENSG00000140262	ENST00000267811: n/a	ENSP00000267811: n/a	INTRONIC	SNPs in gene context

Population genotypes and allele frequencies
This SNP has no allele or genotype frequencies per population.

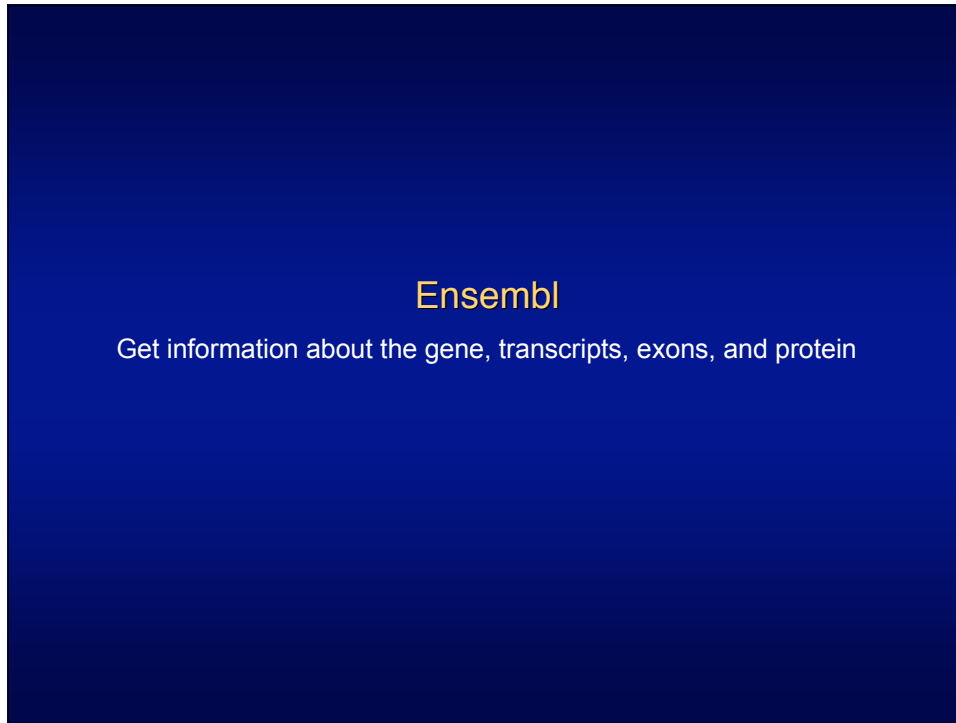
Individual genotypes for SNP rs12915002

SNP Context - 15 54999139

Features Source SNP class SNP type Decorations Export Image size Help

EST trans.
rRNA
tRNA segments
Ensembl trans.
Length 20.00 kb
DNA (contigs)
AC030989.1:221590
Length 20.00 kb
Reverse strand
Ensembl trans.
tRNA segments
rRNA
EST trans.
No nCRNAs in this region
No Ensembl transcripts in this region
No Ensembl transcripts in this region
No nCRNAs in this region
No EST transcripts in this region
8 of the 105 variations in this region have been filtered out by the Source, Class and Type menus.

SNPs
Genotyped SNPs
SNP legend
Intergenic
Synonymous coding
Upstream
Intronic



Ensembl release 48: Homo sapiens Features on Chromosome 15 54996168-55000187

Ensembl ContigView

Ensembl Human ContigView

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

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

Features Comparative DAS Sources Repeats Decorations Export Image size Help

Chromosome 15 tracks:
 - Conservation
 - EST trans
 - Ensembl trans
 - Ensembl Gene (TF12)
 - DNA (contigs)
 - Human MapPath clones
 - SNPs
 - Reg Features
 - Gene legend
 - SNP legend
 - Reg feats legend

Basepair view

© 2008 WTS / EB. Ensembl is available to download for public use - please see the [code licence](#) for details.

Ensembl GeneView

Ensembl release 48 - Dec 2007

Gene: **TCF12** (HGNC Symbol) Synonyms: HEB, HsT17266, HTF4
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 Q92981](#)

Prediction Method: Gene containing both Ensembl predicted transcripts and Havana manual annotation

Transcripts:

Transcript	ENST00000267811	ENSP00000267811	[Transcript info]	[Exon info]	[Peptide info]
TCF12	ENST00000333725	ENSP00000331057	[Transcript info]	[Exon info]	[Peptide info]
NP_996920.1	ENST00000343827	ENSP00000342459	[Transcript info]	[Exon info]	[Peptide info]
NP_996923.1					

Features:

Chr 15: 55.00 Mb to 55.30 Mb

Length: 300,88 kb

Ensembl trans: TCF12 > Common Known ProteinCoding, NP_996920.1 > Common Known ProteinCoding, NP_996923.1 > Common Known ProteinCoding

DNA(contigs): AC010999.6.1.221986 > AC000511.3.186013 > AC000532.1.186050 >

Alignments: This gene can be viewed in genomic alignment with other species
[view genomic alignment with 7 eutherian mammals Pecan](#)
[view genomic alignment with 10 amniota vertebrates Pecan](#)
[view genomic alignment with Rattus norvegicus](#)
[view genomic alignment with Musca domestica](#)
[view genomic alignment with Leiodontia africana](#)
[view genomic alignment with Echinops telfairi](#)
[view genomic alignment with Oryzotagus curvica](#)
[view genomic alignment with Dasypus novemcinctus](#)
[view genomic alignment with Canis familiaris](#)
[view genomic alignment with Pan troglodytes](#)
[view genomic alignment with Gallus gallus](#)
[view genomic alignment with Ornithorynchus anatinus](#)
[view genomic alignment with Bos taurus](#)

Click

Ensembl AlignSliceView

Ensembl release 48 - Dec 2007

Chromosome 15

Overview

Detailed view

Jump to region: 15 : 54998125 - 55368004

Features: Comparative, Repeats, Decorations, Export, Image size

Transcripts:

- ENSEST00000001725 >
- ENSEST00000001727 >
- NP_996920.1 > Common Known ProteinCoding
- TCF12 > Common Known ProteinCoding

DNA(contigs): AC010999.6.1.221986 > AC000511.3.186013 > AC000532.1.186050 >

3cRNA

Canis_familiaris

- ENSICAF700000025746 > Ensembl Novel Protein Coding
- ENSICAF700000025748 > Ensembl Novel Protein Coding

DNA(contigs): ENSICAF700000025746 > ENSICAF700000025748 >

Basepair view

© 2008 WTS / EBI. Ensembl is available to download for public use - please see the [code licence](#) for details.

Ensembl release 48: Homo sapiens Gene report for ENSG00000140262

Ensembl GeneView

Orthologue Prediction

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

Species	Type	Gene identifier
<i>Aedes aegypti</i>	1-to-many	AAEL010226 (Novel Ensembl prediction) [FullContextView] [Align]
daughterless		
[Target: #id: 29; Query: #id: 22]		
<i>Anopheles gambiae</i>	1-to-many	AGAP008814 (Novel Ensembl prediction) [FullContextView] [Align]
No description		
[Target: #id: 31; Query: #id: 22]		
<i>Bos taurus</i>	1-to-1	ENSBTAG0000000586 (AJDN84_BOVIN) [FullContextView] [Align]
HSC142325 protein. [Source: Uniprot/SPTREMBL; Acc: A0JN84]		
[Target: #id: 97; Query: #id: 90]		
<i>Caenorhabditis elegans</i>	1-to-many	MS55.5 (HLH-2) [FullContextView] [Align]
hlh-2 encodes a class II basic helix-loop-helix (BHLH) transcription factor that is the C. elegans ortholog of the mammalian E and Drosophila Daughterless transcriptional activators. HLH-2 activity is required for cell fate specifications occurring during embryonic and larval development that affect such processes as gonadogenesis, male tail formation, and programmed cell death. HLH-2 has been shown to dimerize with at least two C. elegans Achete-acute homologs, LIN-32, a neural-specific protein with which it functions in male tail development and HLH-3, with which it is coexpressed in the nuclei of embryonic neuronal precursors and with which it regulates the transcription of the EGG-1 cell death activator in the NEM sister cells. In gonadogenesis, HLH-2 is required for bestowing proAC competence on the cells that undergo the AC/PU (anterior cell/ventral uterine precursor) cell fate decision, for specification, differentiation, and function of the distal tip cell (DTC) and AC, including transcriptional regulation of the AC-2/3-like ligand in the latter, and for formation of the uterine seam cell (UCA). Genetic analysis also suggests that HLH-2 functions with HLH-14, an additional Achete-acute homolog, to specify the PIV/HSN/PHB neuroblast cell lineage. HLH-2 is expressed in all nuclei of early embryos until the C200-cell stage, when expression becomes increasingly restricted to neuronal cells and their immediate precursors; later expression is detected in, but not limited to, pharyngeal cells, anterior neurons, vulval and uterine muscles, the DTCs, the presumptive and mature AC, the D neuroblast, and other muscles. Comparative analysis of transcriptional and positional reporters indicates that HLH-2 is expressed in both the anterior cell and the ventral uterine (UV) precursor, but that expression in the latter is subject to post-transcriptional down-regulation. HLH-2 accumulation in the presumptive AC is the first detectable difference between the AC and UV precursors during the lateral specification event that distinguishes these two cell fates. [Source: WormBase]		
[Target: #id: 24; Query: #id: 13]		
<i>Canis familiaris</i>	1-to-1	ENSCEAF00000016200 (TCF12) [FullContextView] [Align]
Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SV/SPPROT; Acc: Q99081]		
[Target: #id: 96; Query: #id: 82]		
<i>Cavia porcellus</i>	1-to-1	ENSCEPG00000009153 (TCF12) [FullContextView] [Align]
Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SV/SPPROT; Acc: Q99081]		
[Target: #id: 82; Query: #id: 79]		
<i>Ciona intestinalis</i>	1-to-many	ENSING00000009523 (Q4HN7_CIOIN) [FullContextView] [Align]
Transcription factor protein. [Source: Uniprot/SPTREMBL; Acc: Q4HN7]		
[Target: #id: 24; Query: #id: 21]		
<i>Ciona savignyi</i>	1-to-many	ENSCEAVG00000011705 (Novel Ensembl prediction) [FullContextView] [Align]
No description		
[Target: #id: 22; Query: #id: 20]		
<i>Danio rerio</i>	1-to-1	ENSDBG00000004714 (zgc:85956) [FullContextView] [Align]
transcription factor 12 [Source: RefSeq; Acc: NM_099281]		
[Target: #id: 64; Query: #id: 63]		
<i>Dasypus novemcinctus</i>	1-to-1	ENSNDG00000013864 (TCF12) [FullContextView] [Align]
Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SV/SPPROT; Acc: Q99081]		
[Target: #id: 59; Query: #id: 58]		
<i>Drosophila melanogaster</i>	1-to-many	CG5112 (db) [FullContextView] [Align]
Protein daughterless. [Source: Uniprot/SV/SPPROT; Acc: P11420]		
[Target: #id: 24; Query: #id: 24]		
<i>Echinops telfairi</i>	1-to-1	ENSETEG00000007295 (TCF12) [FullContextView] [Align]
Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SV/SPPROT; Acc: Q99081]		
[Target: #id: 72; Query: #id: 71]		
<i>Erinaceus europaeus</i>	1-to-1	ENSEUEG00000002182 (TCF12) [FullContextView] [Align]
Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SV/SPPROT; Acc: Q99081]		
[Target: #id: 70; Query: #id: 70]		
<i>Felis catus</i>	1-to-1	ENSEFCG00000001867 (TCF12) [FullContextView] [Align]
Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SV/SPPROT; Acc: Q99081]		
[Target: #id: 83; Query: #id: 68]		

Ensembl release 48: Homo sapiens Gene report for ENSG00000140262

Ensembl Human GeneView

Search: e.g. ENSG00000139618, ENSG00000138411

Ensembl release 48 - Dec 2007

Ensembl Gene Report for ENSG00000140262

Gene: **TCF12** (HGNC Symbol) Synonyms: HEB, HsT17266, HTF4 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/SV/SPPROT; Acc: Q99081]

Prediction Method: Gene containing both Ensembl predicted transcripts and Havana manual annotation

Transcripts

Transcript	ENST00000267811	ENSP00000267811	[Transcript info]	[Exon info]	[Peptide info]
NP_996920.1	ENST00000337325	ENSP00000331057	[Transcript info]	[Exon info]	[Peptide info]
NP_996923.1	ENST00000343827	ENSP00000342459	[Transcript info]	[Exon info]	[Peptide info]

Features

Alignments

This gene can be viewed in genomic alignment with other species

- [view genomic alignment with 7 eutherian mammals Pecan](#)
- [view genomic alignment with 10 amniota vertebrates Pecan](#)
- [view genomic alignment with Rattus norvegicus](#)
- [view genomic alignment with Macaca mulatta](#)
- [view genomic alignment with Leontideus africana](#)
- [view genomic alignment with Echinops telfairi](#)
- [view genomic alignment with Oryzodagus cuniculus](#)
- [view genomic alignment with Dasypus novemcinctus](#)
- [view genomic alignment with Canis familiaris](#)
- [view genomic alignment with Pan troglodytes](#)
- [view genomic alignment with Gallus gallus](#)
- [view genomic alignment with Ornithorynchus anatinus](#)
- [view genomic alignment with Bos taurus](#)

Ensembl release 48: Homo sapiens Gene report for ENSG00000140262

Ensembl GeneView

Search: >

Ensembl release 48 - Dec 2007

HOME - BLAST - BIOMART - SITEMAP - HELP

Your Ensembl

- Login or Register
- About User Accounts

ENSG00000140262

- Gene information
- Gene regulation info.
- Genomic sequence
- Genomic sequence alignment
- Resequencing alignment
- Gene splice site image
- Gene tree info.
- Gene variation info.
- LD history
- Compare SNPs in track
- Transcript info
- Exon information
- Protein information
- Export gene data

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

- View of Chromosome 15
- Graphical view
- Graphical overview
- Export from region...
- Export Gene info in region
- Export SNP info in region
- Export Vega info in region

Ensembl Archive

- View previous release of page in Archive!
- Stable Archive! link for this page

Ensembl Gene Report for ENSG00000140262

Gene: **TCF12** (HGNC Symbol) Synonyms: HEB, HsT17266, HTF4
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](#), [1,221,986](#).

Description: Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Unikent/CS/ISSPOT_09001](#)

Prediction Method: Gene containing both Ensembl predicted transcripts and Havana manual annotation

Transcripts

TCF12	ENST00000267811	ENSP00000267811	[Transcript info]	[Exon info]	[Peptide info]
NP_996920.1	ENST00000333725	ENSP00000331057	[Transcript info]	[Exon info]	[Peptide info]
NP_996923.1	ENST00000343827	ENSP00000342459	[Transcript info]	[Exon info]	[Peptide info]

Features

Alignments

This gene can be viewed in genomic alignment with other species

- [view genomic alignment with 7 eutherian mammals Pecan](#)
- [view genomic alignment with 10 amniota vertebrates Pecan](#)
- [view genomic alignment with Rattus norvegicus](#)
- [view genomic alignment with Musca musca](#)
- [view genomic alignment with Leontidea africana](#)
- [view genomic alignment with Echinops telfairi](#)
- [view genomic alignment with Oryctolagus cuniculus](#)
- [view genomic alignment with Dasypus novemcinctus](#)
- [view genomic alignment with Canis familiaris](#)
- [view genomic alignment with Pan troglodytes](#)
- [view genomic alignment with Gallus gallus](#)
- [view genomic alignment with Ornithorynchus anatinus](#)
- [view genomic alignment with Bos taurus](#)

Ensembl release 48: Homo sapiens Exon Report for ENST00000267811

Ensembl ExonView

Back Forward Reload Stop Home http://www.ensembl.org/Homo_sapiens/exonview?db=core;transcript=ENST00000267811

Exon	Transcript	Start	End	Strand	Phase	Length	Score	
18	ENSE00001126365	15	1	55,352,520	55,352,752	2	1	233
Intron 18-19								9,182
19	ENSE00001103528	15	1	55,361,935	55,362,088	1	-	154
Intron 19-20								3,558
20	ENSE00001405489	15	1	55,365,647	55,368,004	-	-	2,358

3' downstream sequence

aaatagtaggttttacattctctcttcaataaataagatttttg.....

Supporting Evidence

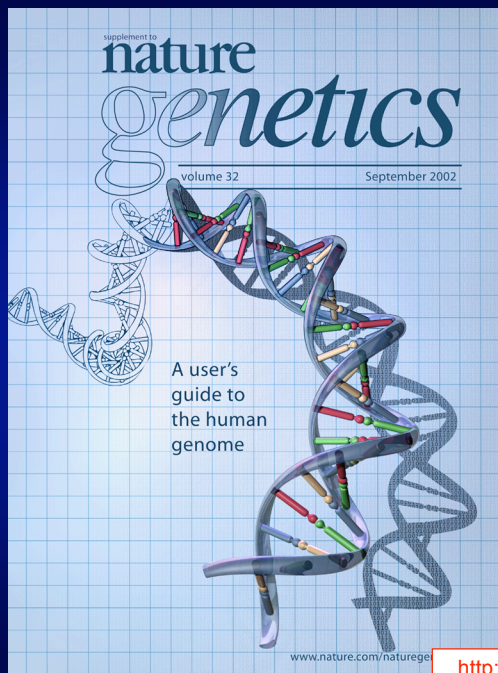
The supporting evidence below consists of the sequence matches on which the exon predictions were based and are sorted by alignment score.

There are a large number of supporting evidence hits for this transcript. Only the top ten hits have been shown. [Click to view all 23 supporting evidence hits.](#)

Score	Source	Sequence
100	NM_207036.1	gH4937007[Ensembl_200708.1] Homo sapiens transcription factor 12 (HTF4, helix-loop-helix transcription factor 4) (TCF12), transcript variant 1, mRNA
99	NM_002005.3	gH4937007[Ensembl_200205.3] Homo sapiens transcription factor 12 (HTF4, helix-loop-helix transcription factor 4) (TCF12), transcript variant 3, mRNA
97	BC051769.2	BC051769.2 Homo sapiens cDNA clone IMAGE5752603, containing transcribed mRNA
95	BC050596.1	BC050596.1 Homo sapiens transcription factor 12 (HTF4, helix-loop-helix transcription factor 4), mRNA in DNA clone IMC-5780 (IMAGE-578759), complete cds
94	BC043924.1	BC043924.1 Homo sapiens transcription factor 12 (HTF4, helix-loop-helix transcription factor 4), mRNA in DNA clone IMC-5780 (IMAGE-578759), complete cds

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 Help & Tutorials
<http://www.ensembl.org/info/using/website>



http://www.nature.com/ng/focusup_index.html

Current Protocols in Bioinformatics

<h3>The UCSC Genome Browser</h3> <p>The rapid progress of public sequencing and mapping efforts on vertebrate genomes has increased the demand for tools that offer quick and easy access to genomic data and facilitate comparative data analysis. The University of California, Santa Cruz (UCSC) Genome Bioinformatics Web site at http://genome.ucsc.edu provides a variety of genome analysis tools, most notably the UCSC Genome Browser (Hinrichs et al., 2006), a graphical tool for viewing genomic data and a collection of aligned annotation "tracks." Another tool, the Table Browser—supplies convenient access to the MySQL database underlying the Genome Browser annotation system. The custom annotation tracks feature that enables users to upload and compare their own tracks.</p> <p>The main protocol of this unit (see Basic Protocol) describes how to perform a simple text-based search of genome annotations to view the genome sequence, zoom in and out, and change the tracks shown. It also describes some of NCBI's tools which are provided as links from the Map Viewer. The Alternate Protocol 1 shows how to perform a BLAST search against the human genome. Alternate Protocol 2 shows how to retrieve a list of all genes between two STS markers. Alternate Protocol 3 shows how to find all annotated members of a gene family.</p>	UNIT 1.4
<h3>Using the NCBI Map Viewer to Browse Genomic Sequence Data</h3> <p>The NCBI Map Viewer is an interface to a large, integrated set of genomic data, including sequence, cytogenetic, genetic linkage, and radiation hybrid maps, as well as the assembled and annotated genomic sequence itself. Along with the UCSC Genome Browser (UNIT 1.4) and Ensembl (UNIT 1.15), it is one of the primary Web sites from which genomic sequence data can be accessed.</p> <p>This unit includes an introduction to the Map Viewer (see Basic Protocol), which describes how to perform a simple text-based search of genome annotations to view the genome sequence, zoom in and out, and change the tracks shown. It also describes some of NCBI's tools which are provided as links from the Map Viewer. The Alternate Protocol 1 shows how to perform a BLAST search against the human genome. Alternate Protocol 2 shows how to retrieve a list of all genes between two STS markers. Alternate Protocol 3 shows how to find all annotated members of a gene family.</p> <p>NCBI provides Map Viewers for eleven vertebrates, six invertebrate plants, and fourteen fungi. Although the data themselves vary, the basic navigation principles are the same. The Basic Protocols 1 and 2 are illustrated with examples from the human genome. Protocol 3 uses the mouse genome.</p>	UNIT 1.5
<h3>Using the Ensembl Genome Server to Browse Genomic Sequence Data</h3> <p>The Ensembl project presents the latest sequence assembly of the human genome and provides automatic annotation of that sequence, including gene, transcript, and protein predictions. The annotation is integrated with external data sources, making Ensembl a valuable starting and reference point for any work in human biology or medicine that utilizes genetic information.</p> <p>A central element of the Ensembl project is openness: all data are freely available and all the computer code used to analyze and present the data is freely available as well. More information on the Ensembl gene prediction and annotation system, and on additional ways of accessing the data, is provided in the Commentary.</p> <p>This unit explains how to access and use the human sequence (although these instructions would be applicable to any of the species available in the browser) and its annotation via the Ensembl Web site. The Web site is an advanced interactive service, providing a range of views that present different aspects of the data. The Ensembl human home page (http://www.ensembl.org/Homo.sapiens) provides access to the data in several different ways, including text searches, clickable chromosomes, and sequence similarity searching in BLASTView, as well as by using the BioMart data warehouse or by simply entering chromosome coordinates.</p>	UNIT 1.15

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

Current Topics in Genome Analysis

Next Lecture:

Evolutionary Analysis

Fiona Brinkman, Ph.D.
Simon Fraser University