


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/>

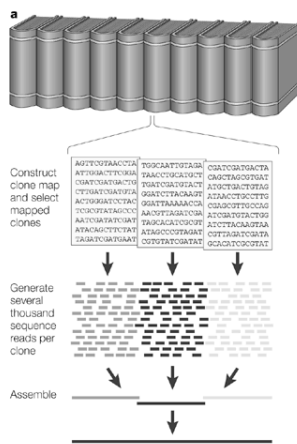
Ensembl
<http://www.ensembl.org>

Types of data integrated in genome browsers

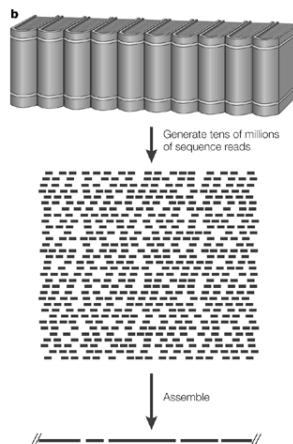
- Same starting material for all genome browsers: genomic sequence
- Annotations calculated independently by each genome browser
 - Genes
 - RefSeq mRNAs (non-redundant)
 - GenBank mRNAs (redundant)
 - Gene predictions
 - SNPs
 - Homologous sequences from other organisms
 - STSs

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/>
 - pre!Ensembl: <http://pre.ensembl.org/>
 - UCSC and Ensembl provide archive of all genome assemblies and annotations; NCBI provides only limited archive
- IF YOU ARE COMPARING DATA FROM DIFFERENT GENOME BROWSERS, MAKE SURE YOU ARE LOOKING AT THE SAME VERSION OF THE ASSEMBLY

Genome Assembly Versions

	Same assembly?	UCSC	NCBI	Ensembl
Human	Yes	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

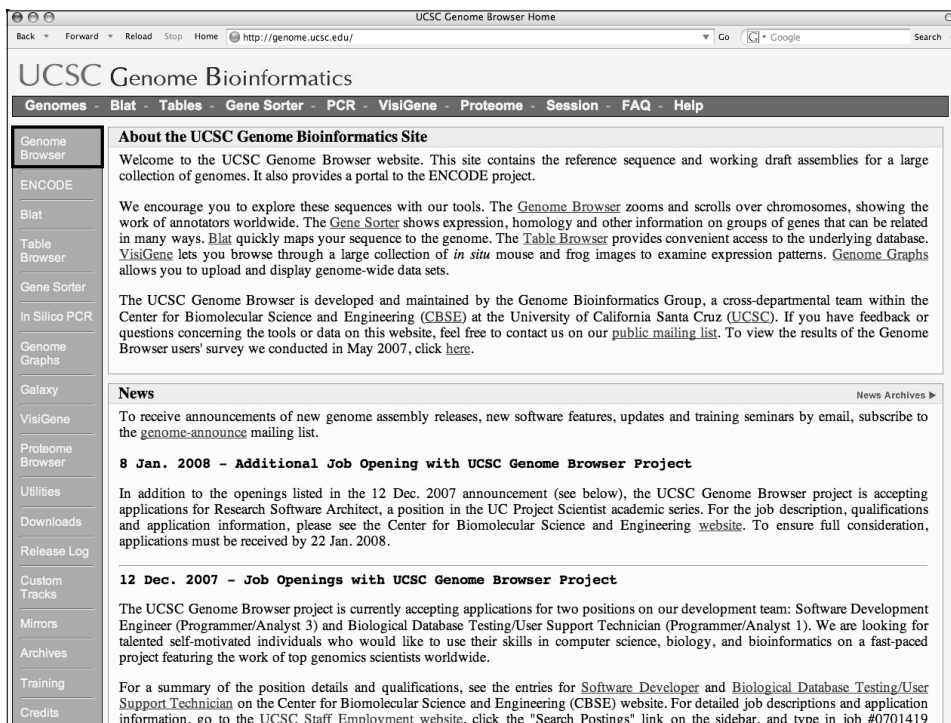
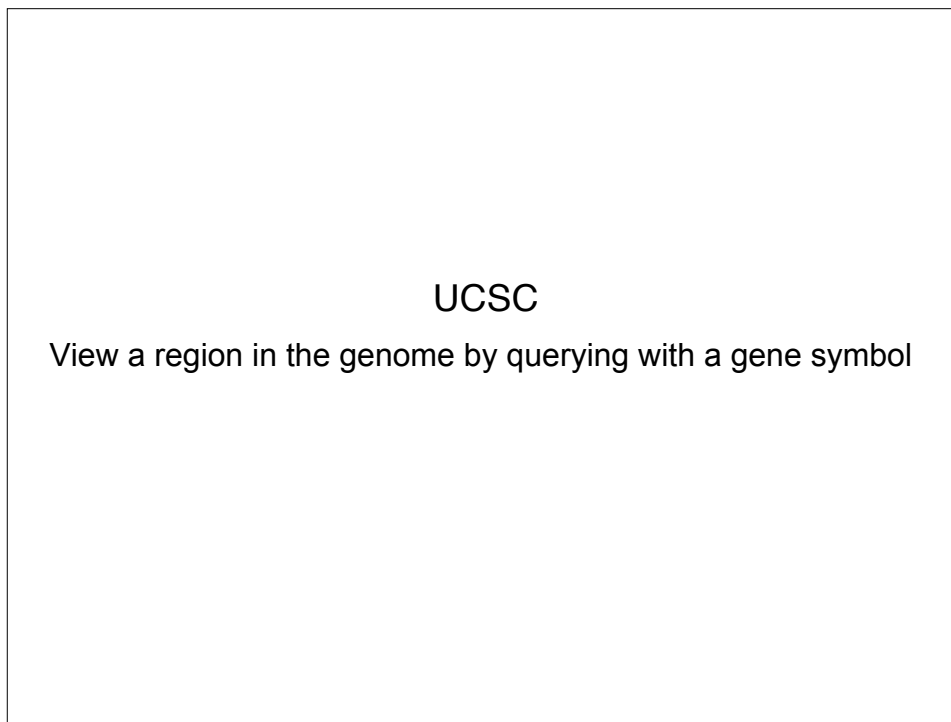
NCBI Reference Sequences (RefSeqs)

- 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>

<pre> 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 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Theria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo. REFERENCE 1 (bases 1 to 1793) AUTHORS 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 17933480 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:5016088. 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. COMPLETE: complete on the 3' end. COMPLETE: complete on the 3' end. </pre>	<p>Beta actin mRNA RefSeq</p>
<pre> CDS 74..1201 /gene="ACTB" /locus_tag="actin filament; cytoskeleton; TIF60 histone acetyltransferase complex [PMID 1096108]" /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:60" /db_xref="HGN:117" /db_xref="HPRD:00032" /db_xref="MIM:103610" /translation="MDDIALLVVNGSGMCKAGFAGDDAPRAVFPFSIVGRPHIQGVN VMGQKDYVGDAGSKRGLTLKYPTEGIVNMDMEKIMHFFYNEIRVAPENP VLTGAPLRFNANRKNFYKMFPTFPFANVAIVAVLSLYAGSTTIVWGGDVE HTPVLYSGVALPHAILRLDLACRDLTDYLMKILTERGVSPFTTAREIRVRIKELCY VALDFQEMATAASSSLEKSYELPDQGVITIGNERKCPALFPSPFLGMSGCLHE FTNSHRCQVPIKSLVAFVIGGFTWPTFADRMKSTALASPTMKIAPPE RKYSVWIGSILASLSTFQGMWISKQEDYESGSPSIVHRKCF" </pre>	<pre> ORIGIN 1 cggctcgcgc ccgcgagcac agagctctgc ctttgcgat cgcgcgccgc tccacaccgc 61 ccgcacgctc accatggatg atgatatcgc cgcgctctgc gtgcacaacg gctccggcat 121 gtgcaagggc ggtcttcgcy gcgacgatgc cccccggccc gctctcccc ccatgctggg</pre>



Human (Homo sapiens) Genome Browser Gateway

Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help

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

clade genome assembly position or search term image width

Vertebrate Human Mar. 2006 ADAM2 820 submit **click**

Chimp
Rhesus
Mouse
Rat
Cat
Dog
Horse

Human Mar. 2006 (hg18)
Cow
Opossum
Platypus
Chicken
Lizard
X. tropicalis
Zebrafish
Tetraodon
Fugu
Stickleback
Medaka

Reset the browser user interface settings to their defaults.
tracks | configure tracks and display | clear position

About the Human Mar. 2006 (hg18)
The March 2006 human reference sequence (36.1) was produced by the International Human Genome Sequencing Consortium.

Sample position queries
A genomic position can be specified by the chromosomal coordinate range, or keyword, human genome. See the [User's Guide](#) for details.

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

D1683046 Displays region around STS marker D1683046 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

Human ADAM2 - UCSC Genome Browser v1.74

Back Forward Reload Stop Home http://genome.ucsc.edu/cgi-bin/hgTracks?hgid=103237419&clade=vertebrate&org=Human&db=hg18 Go Google Search

UCSC Genes

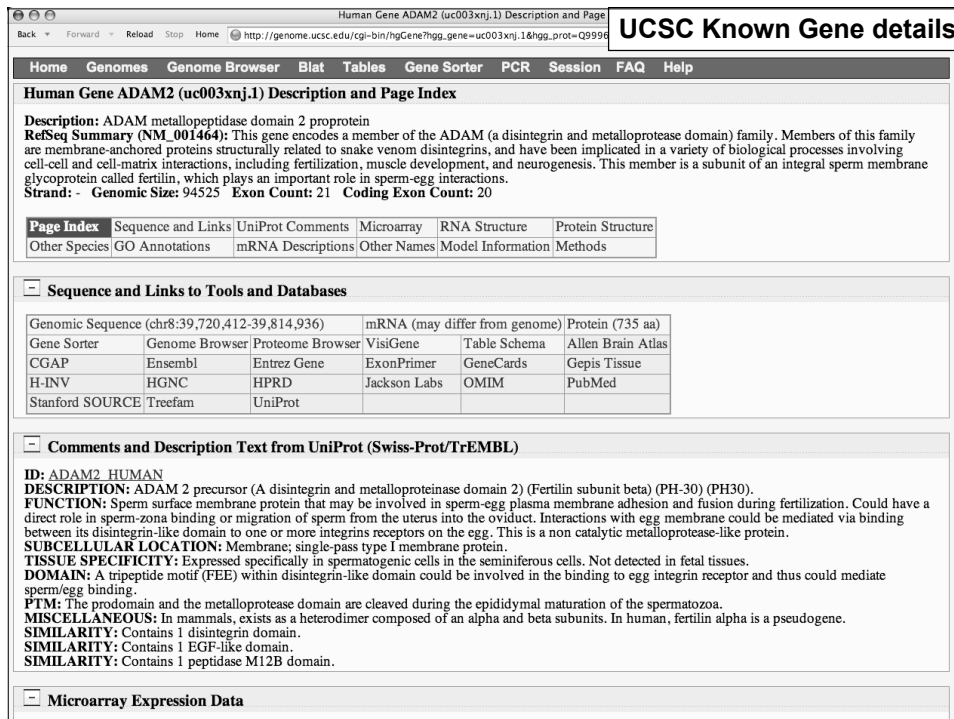
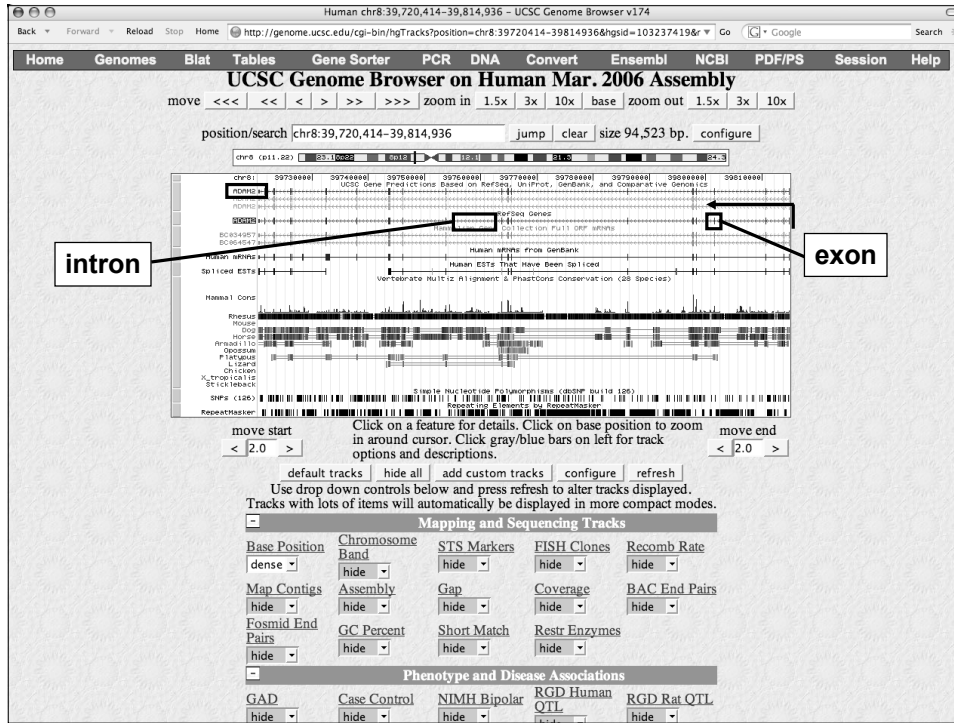
ADAM2 (uc003xnl.1) at chr8:39720412-39814936 - ADAM metalloproteinase domain 2 preprotein
ADAM2 (uc003xnl.1) at chr8:39720412-39814936 - ADAM metalloproteinase domain 2 preprotein
ADAM2 (uc003xnl.1) at chr8:39720412-39814936 - ADAM metalloproteinase domain 2 preprotein
ADAM2B (uc003kdv.1) at chr8:24207525-24268670 - ADAM metalloproteinase domain 28 isoform 1
ADAM2B (uc003kdv.1) at chr8:24207525-24268670 - ADAM metalloproteinase domain 28 isoform 3
ADAM22 (uc003ajp.1) at chr7:87401638-87664385 - ADAM metalloproteinase domain 22 isoform 4
ADAM22 (uc003ajp.1) at chr7:87401638-87664385 - ADAM metalloproteinase domain 22 isoform 3
ADAM22 (uc003ajn.1) at chr7:87401638-87664385 - ADAM metalloproteinase domain 22 isoform 1
ADAM22 (uc003ajm.1) at chr7:87401638-87664385 - ADAM metalloproteinase domain 22 isoform 2
ADAM22 (uc003ajk.1) at chr7:87401638-87664385 - ADAM metalloproteinase domain 22 isoform 5
ADAM22 (uc003ajl.1) at chr7:87401638-87664385 - ADAM metalloproteinase domain 22 isoform 2
ADAM22 (uc003ajh.1) at chr7:87401638-87664385 - ADAM metalloproteinase domain 22 isoform 4
ADAM29 (uc003iue.1) at chr4:17607634-176135906 - ADAM metalloproteinase domain 29 preprotein
ADAM29 (uc003iud.1) at chr4:17607634-176135906 - ADAM metalloproteinase domain 29 preprotein
ADAM29 (uc003iue.1) at chr4:17607634-176135906 - ADAM metalloproteinase domain 29 preprotein
ADAM23 (uc002vbg.1) at chr2:207016613-207190924 - ADAM metalloproteinase domain 23 preprotein
ADAM20 (uc001xme.1) at chr14:70058831-70071485 - ADAM metalloproteinase domain 20 preprotein
ADAM21 (uc001xmd.1) at chr14:69993970-69996375 - ADAM metalloproteinase domain 21 preprotein

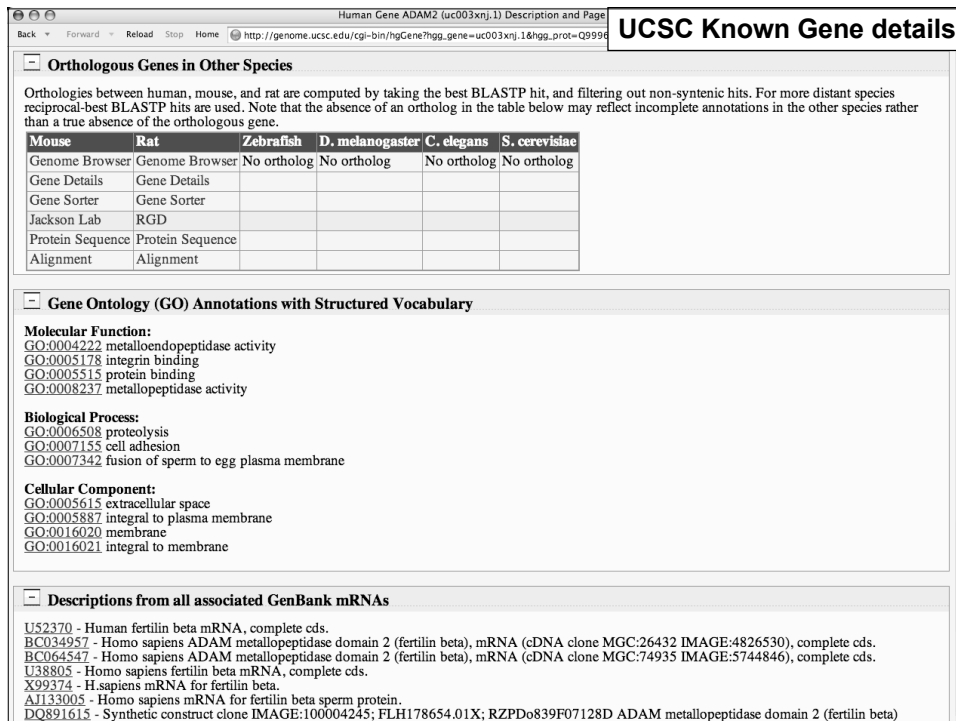
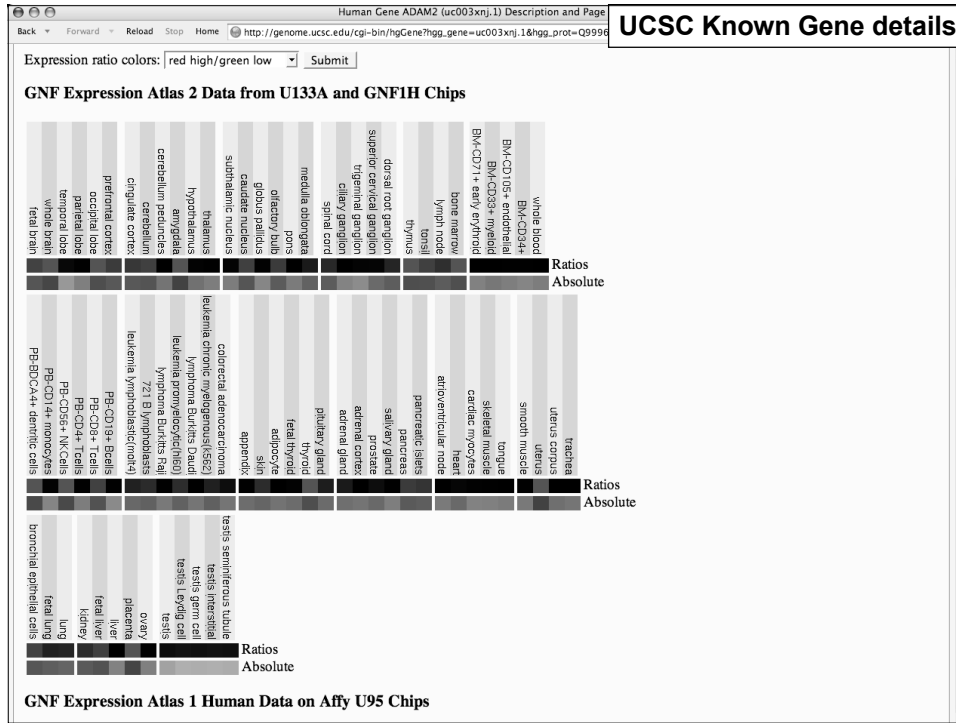
RefSeq Genes

ADAM2 at chr8:39720412-39814936 - (NM_001464) ADAM metalloproteinase domain 2 preprotein
ADAM20 at chr14:70058831-70071485 - (NM_003814) ADAM metalloproteinase domain 20 preprotein
ADAM21 at chr14:69993970-69996375 - (NM_003813) ADAM metalloproteinase domain 21 preprotein
ADAM22 at chr7:87401638-87664385 - (NM_004194) ADAM metalloproteinase domain 22 isoform 4
ADAM22 at chr7:87401638-87664385 - (NM_021721) ADAM metalloproteinase domain 22 isoform 5
ADAM22 at chr7:87401638-87664385 - (NM_021722) ADAM metalloproteinase domain 22 isoform 2
ADAM22 at chr7:87401638-87664385 - (NM_021723) ADAM metalloproteinase domain 22 isoform 1
ADAM22 at chr7:87401638-87664385 - (NM_016351) ADAM metalloproteinase domain 22 isoform 3
ADAM23 at chr2:207016613-207190924 - (NM_003812) ADAM metalloproteinase domain 23 preprotein
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 chr4:17607634-176135906 - (NM_014269) ADAM metalloproteinase domain 29 preprotein

Non-Human RefSeq Genes

ADAM2 at chr8:39723140-39814873 - (NM_001082677) ADAM metalloproteinase domain 2 (fertilin beta)
ADAM2 at chr8:39723140-39814873 - (NM_212957) ADAM metalloproteinase domain 2
ADAM2 at chr8:39721536-39813888 - (NM_174228) ADAM metalloproteinase domain 2
Adam2 at chr8:39723147-39801569 - (NM_009618) a disintegrin and metalloproteinase domain 2
Adam2 at chr8:39723147-39813877 - (NM_020077) a disintegrin and metalloproteinase domain 2
Adam21 at chr14:69998947-69996354 - (NM_020330) a disintegrin and metalloproteinase domain 21
Adam21 at chr14:69781958-70061197 - (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
Adam22 at chr7:87401747-87670140 - (NM_001098225) a disintegrin and metalloproteinase domain 22
Adam22 at chr2:207016503-207194450 - (NM_011780) a disintegrin and metalloproteinase domain 23
Adam23 at chr2:207016503-207194450 - (NM_001029899) a disintegrin and metalloproteinase domain 23
Adam24 at chr1:188904672-188906248 - (NM_010086) a disintegrin and metalloproteinase domain 24
Adam25 at chr1:188904411-188906248 - (NM_011781) a disintegrin and metalloproteinase 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
- Genomic Sequence from assembly

Click

UCSC RefSeq Gene details

Genomic Sequence Near Gene

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

Get Genomic Sequence Near Gene

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

Sequence Retrieval Region Options:

Promoter/Upstream by 1000 bases
 5' UTR Exons
 CDS Exons
 3' UTR Exons
 Introns
 Downstream by 1000 bases
 One FASTA record per gene.
 One FASTA record per region (exon, intron, etc.) with 0 extra bases upstream (5') and 0 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
ggaagtatctaccaaacacatccctgtgatccgacaactcactctagaa
atatacaacagtagaaccctacttatcacocaaagcctgtagaaga
atgttatagctaaatatttttaaatagctggaacataaacacaaca
aatatcctaaacagtaaaatggaacacaagaaggtgttatattatga
attgtatatacaccaatgaggataaacagaactatgcttgtataga
accttaaacatcctattaaaagaaccccaacatgaaagtagtgggt
gattgcttctactgcaaaaagtcaaaaacagacaacaaacgaatcttgg
ttgtttagaagtcaggtgtgaggtggaatctggggatttgggtgggt
ctttttcaatctttcaacgggtgactagtttaaggttttttttcaac
ttgaatattatgacctggaactatgattattataacttttttctc
gttttttttctctctttttttttttttttttttttttttttttttt
ttctgctctcaccacaggtggagtgcaaggtttctctgctcaac
tgcaacctgctcctagggtcaagcagattctctgctcagctctccg
agttagctgggattcaggcaaccgcaacatgctgctaatttttttt
gtatttttagtaacagcgggtttcaocattgttgcggaactggtctcga
actcctgatcaggttttatatatattcaattgaaattttacttaagaagt
gtttataaatctctctgttctcagctgttgaagtgatttggttgttgc
tgttgccatttaaggatcaacgtcccgtaggtgtgttctctcgaagag
acagggtcaggagctgcagcgttccacagcgcacacacccaactcag
cccacctgggctctcccaagcctcctctccagcgtgctggccggg
```


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,...

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)

Rhesus Mouse Dog Rhesus Macaque

Primate 110 Genomes

Chromosome X Y Chromosomes

SNPs (126)

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

.....

Expression and Regulation

Affy All Exon	Affy HuEx 1.0	GENE 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 TAFI...	ORegAnno	Affy Txn...
hide	hide	hide	hide	hide
hide dense squish pack full	7X Reg Potential	Uppsala ChIP...		

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 (b1,28-31,21)

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

RefSeq Genes

Human mRNAs from GenBank

Human ESTs That Have Been Spliced

MMR Conserved Transcription Factor Binding Sites

Vertebrate Multiz Alignment & PhastCons Conservation (20 Species)

Rhesus Mouse Dog Rhesus Macaque

Primate 110 Genomes

Chromosome X Y Chromosomes

SNPs (126)

RepeatMasker

move start < 2.0 > move end < 2.0 >

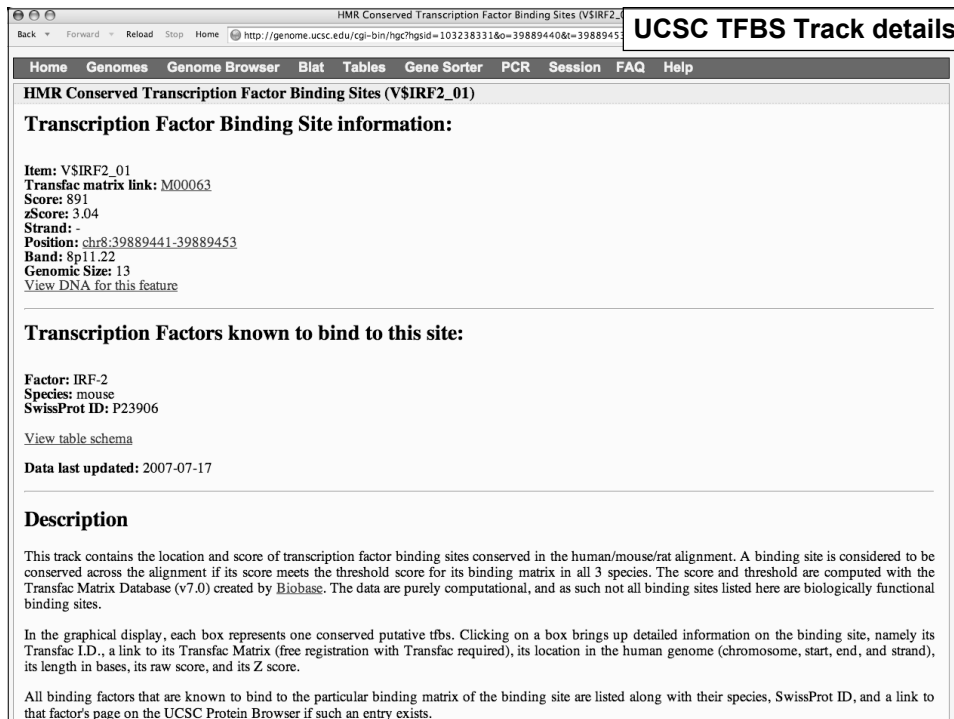
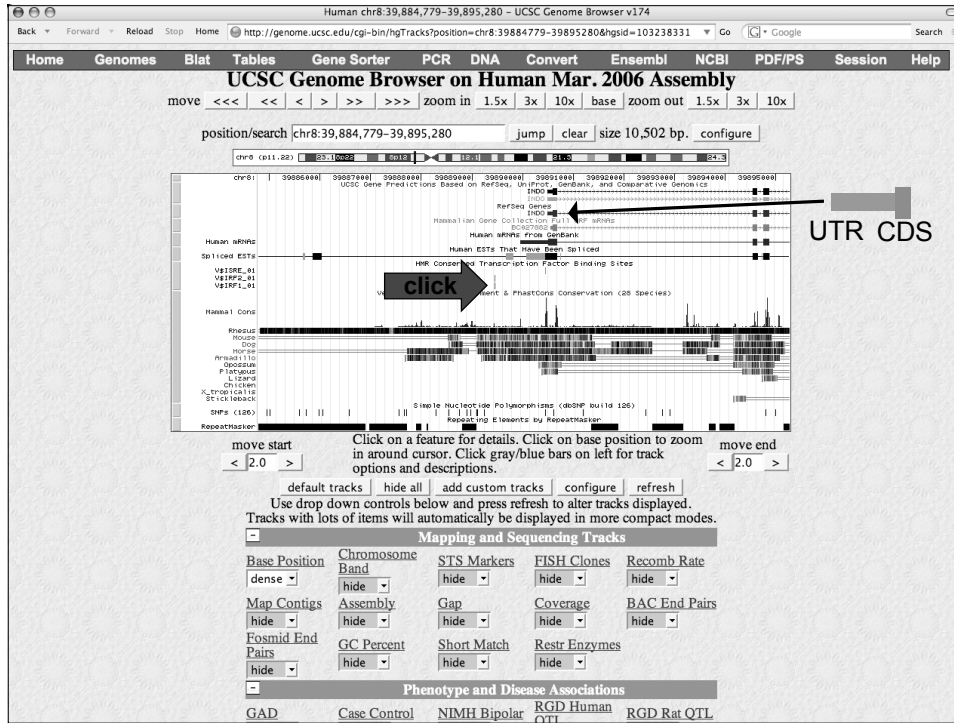
Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

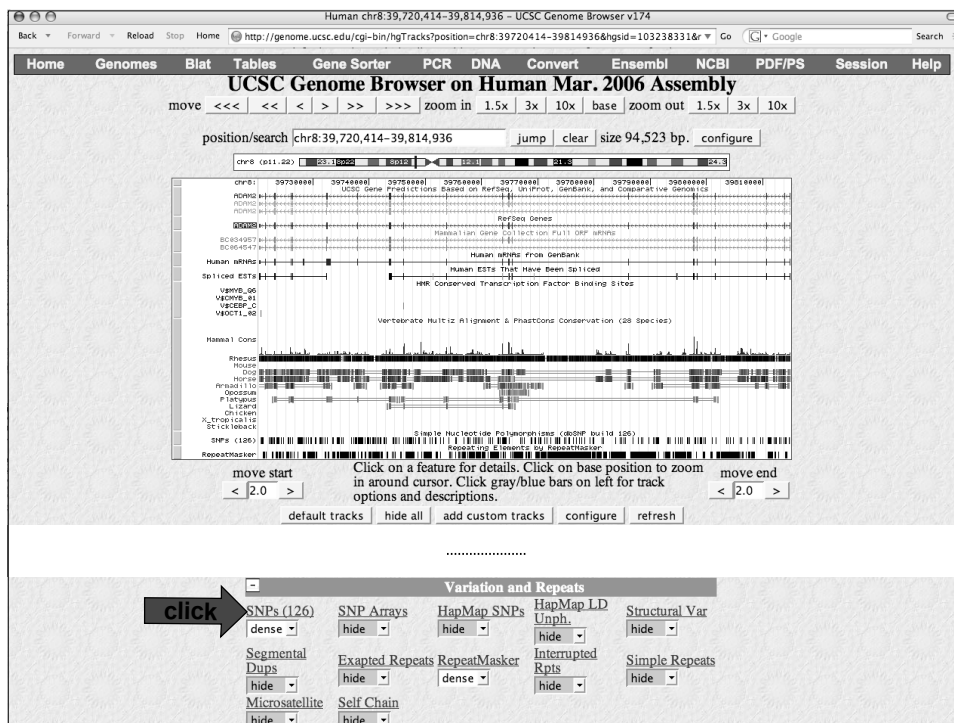
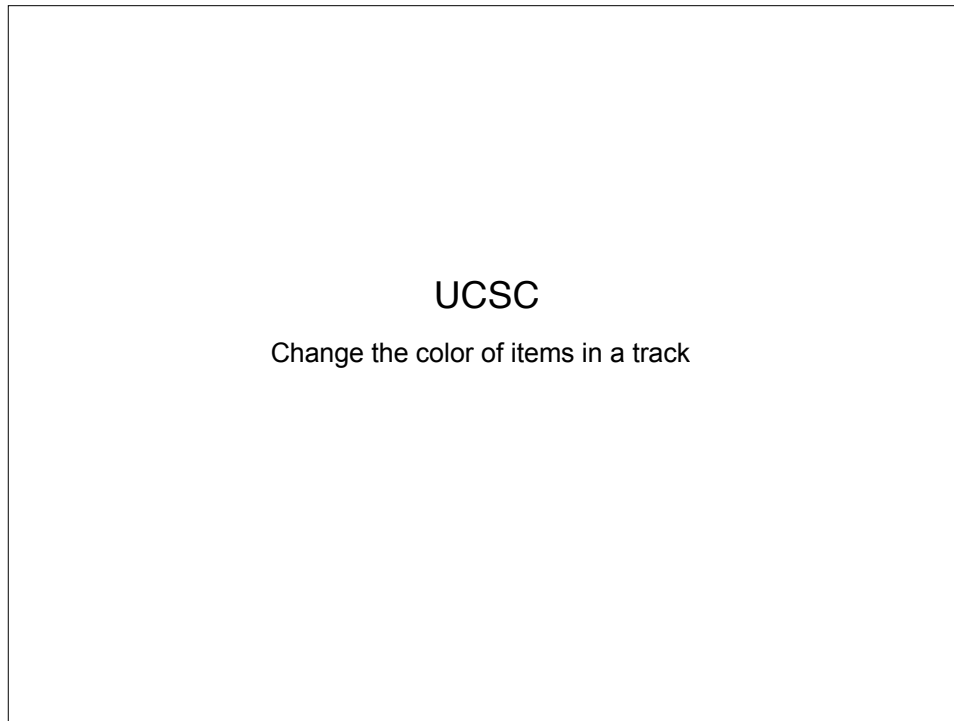
default tracks hide all add custom tracks configure refresh

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





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

http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr8:39720414-39814936&hgslid=103238331&r

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) 39,720,414 39,814,936

chr8: 39738888 39748888 39758888 39768888 39778888 39788888 39798888 39808888 39818888

UCSC Gene Predictions Based on RefSeq, Ensembl, and Comparative Genomics

RefSeq Genes

Human Gene Collection Full RefSeq entries

Human mRNAs from GenBank

Human ESTs that have been spliced

Spliced ESTs

Human Conserved Transcription Factor Binding Sites

Vertebrate Multiple Alignment & PhastCons Conservation (28 Species)

Human Cons

Rhesus Mouse

Dog

Primate 10

Opposum

Platypus

Chicken

Human

SNPs (126)

RepeatMasker

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 start < 2.0 > move end < 2.0 >

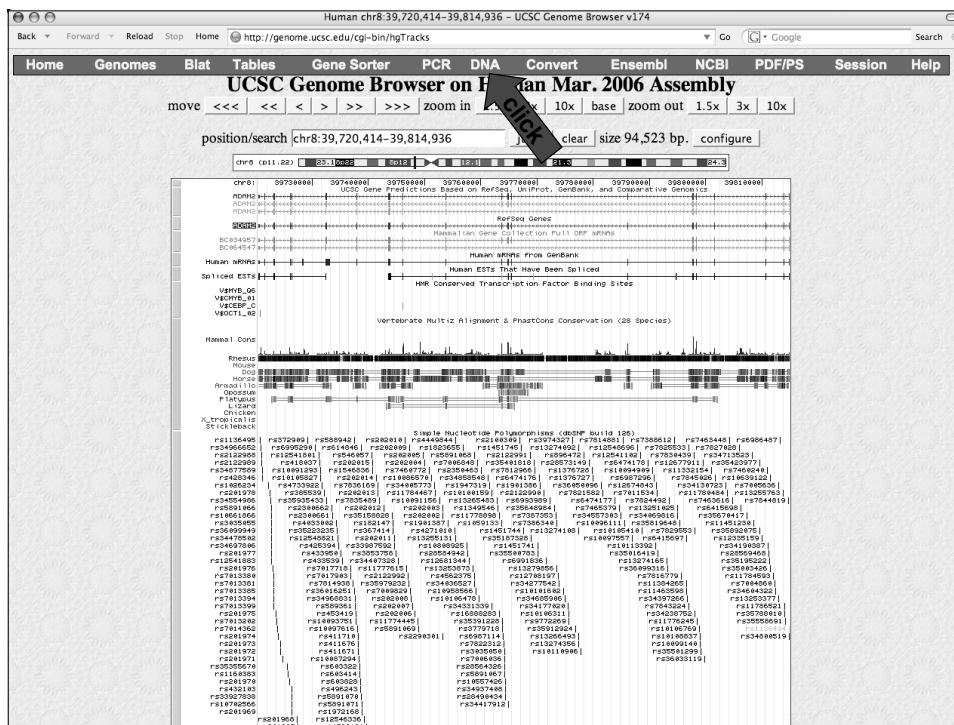
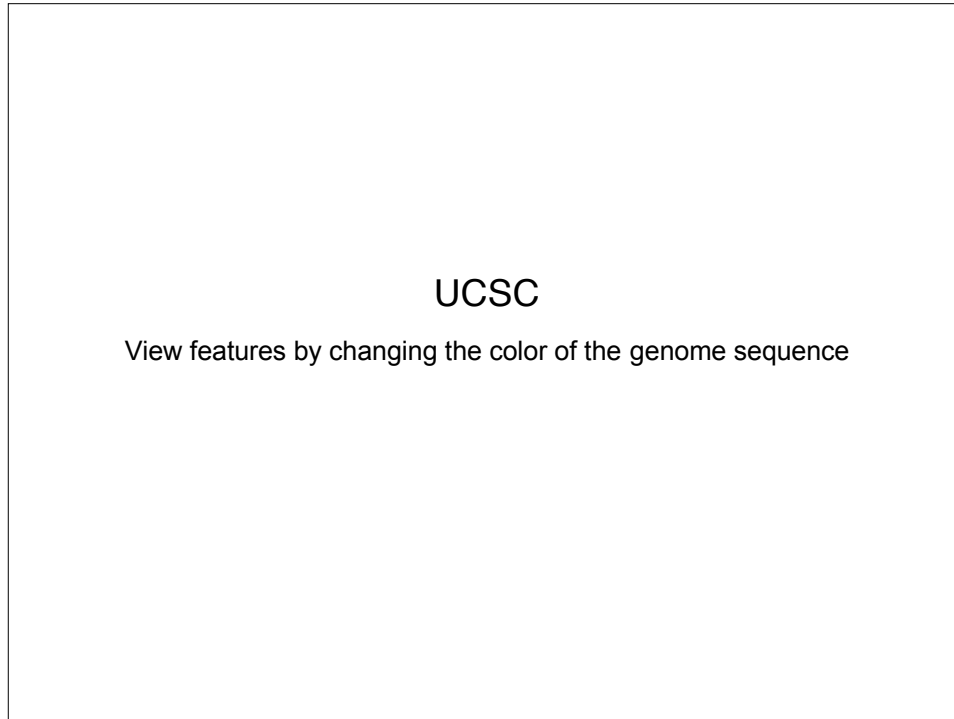
default tracks hide all add custom tracks configure refresh

.....

click →

Variation and Repeats

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



Entrez Protein

NCBI Entrez Protein

Search Protein for np_001455

Display Summary Show 20 Sort by Relevance Send to

1: NP_001455. Reports ADAM metallopeptidase domain 2 proprotein [Homo sapiens]

```
>gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
MWRVLLSGLGGLRMDNSFSLPVQITVPEKIRSIKKGIESQASYKIVIRGKPYTVNLMQKNFLPHNF
RVYSYSGTQIMKPLDQDFONPCHYQYIROYPKSVVMVSPCTLROGLQFNYSYGLPESLSSVGFEEVT
YQYKRRADVSLYNERDIESRDLSEFLQSVPODFAKYIEMHVIVEKGLYNMGSDDTVVAKVFLQIG
L7NAIFVSPNITILLSSLELWIDENKIAT7GEANELLH7FLRWKTSYL7L7RPHDVAFL7LVYREKSNV7GA
7FQGMKCDANYAGGV7LHPT7ISL7SLAY7L7AQL7LSMG7IT7DD7IN7K7C7G7S7G7A7C7M7N7P7E7A7L7H7P7S7G7V7L
7F8N7S7F7D7P7A7H7I7E7K7R7K7M7C7R7P7S7F7E7C7D7L7P7E7C7N7G7S7A7S7C7P7E7N7H7V7O7T7G7H7P7C7L7N7O7W7I7C7D7G7V7C7M
7F7R7A7G7S7N7C7A7D7G7P7C7C7E7N7C7L7F7M7S7K7E7M7C7R7P7S7F7E7C7D7L7P7E7C7N7G7S7A7S7C7P7E7N7H7V7O7T7G7H7P7C7L7N7O7W7I7C7D7G7V7C7M
7S7D7K7C7T7D7F7G7R7E7V7F7E7G7P7S7E7Y7S7H7L7N7S7K7T7D7V7S7N7G7I7S7D7S7Y7T7C7E7A7D7N7L7C7Q7K7L7C7K7Y7V7G7F7L7Q7I7P7R7A
7T7I7Y7A7N7I7S7G7L7C7I7A7V7E7F7A7S7D7H7A7D7S7K7M7I7K7D7T7S7C7E7N7K7V7C7R7N7C7V7S7S7L7O7C7T7T7D7K7N7D7R7C7V7C7N7K
7K7C7H7C7S7A7S7Y7L7P7D7C7V7S7D7L7W7G7S7I7D7S7N7F7P7V7A7I7P7A7L7P7E7R7Y7I7E7N7I7Y7H7S7K7P7M7W7P7F7L7I7P7F7I7F7C
7V7L7A7L7M7V7N7P7Q7R7K7R7T7E7D7S7D7Q7E7S7E7P7K7G
```

Chicken BLAT Search

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 metallopeptidase domain 2 proprotein [Homo sapiens]
MWRVLLSGLGGLRMDNSFSLPVQITVPEKIRSIKKGIESQASYKIVIRGKPYTVNLMQKNFLPHNF
RVYSYSGTQIMKPLDQDFONPCHYQYIROYPKSVVMVSPCTLROGLQFNYSYGLPESLSSVGFEEVT
YQYKRRADVSLYNERDIESRDLSEFLQSVPODFAKYIEMHVIVEKGLYNMGSDDTVVAKVFLQIG
L7NAIFVSPNITILLSSLELWIDENKIAT7GEANELLH7FLRWKTSYL7L7RPHDVAFL7LVYREKSNV7GA
7FQGMKCDANYAGGV7LHPT7ISL7SLAY7L7AQL7LSMG7IT7DD7IN7K7C7G7S7G7A7C7M7N7P7E7A7L7H7P7S7G7V7L
7F8N7S7F7D7P7A7H7I7E7K7R7K7M7C7R7P7S7F7E7C7D7L7P7E7C7N7G7S7A7S7C7P7E7N7H7V7O7T7G7H7P7C7L7N7O7W7I7C7D7G7V7C7M
7F7R7A7G7S7N7C7A7D7G7P7C7C7E7N7C7L7F7M7S7K7E7M7C7R7P7S7F7E7C7D7L7P7E7C7N7G7S7A7S7C7P7E7N7H7V7O7T7G7H7P7C7L7N7O7W7I7C7D7G7V7C7M
7S7D7K7C7T7D7F7G7R7E7V7F7E7G7P7S7E7Y7S7H7L7N7S7K7T7D7V7S7N7G7I7S7D7S7Y7T7C7E7A7D7N7L7C7Q7K7L7C7K7Y7V7G7F7L7Q7I7P7R7A
7T7I7Y7A7N7I7S7G7L7C7I7A7V7E7F7A7S7D7H7A7D7S7K7M7I7K7D7T7S7C7E7N7K7V7C7R7N7C7V7S7S7L7O7C7T7T7D7K7N7D7R7C7V7C7N7K
7K7C7H7C7S7A7S7Y7L7P7D7C7V7S7D7L7W7G7S7I7D7S7N7F7P7V7A7I7P7A7L7P7E7R7Y7I7E7N7I7Y7H7S7K7P7M7W7P7F7L7I7P7F7I7F7C
7V7L7A7L7M7V7N7P7Q7R7K7R7T7E7D7S7D7Q7E7S7E7P7K7G
```

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.

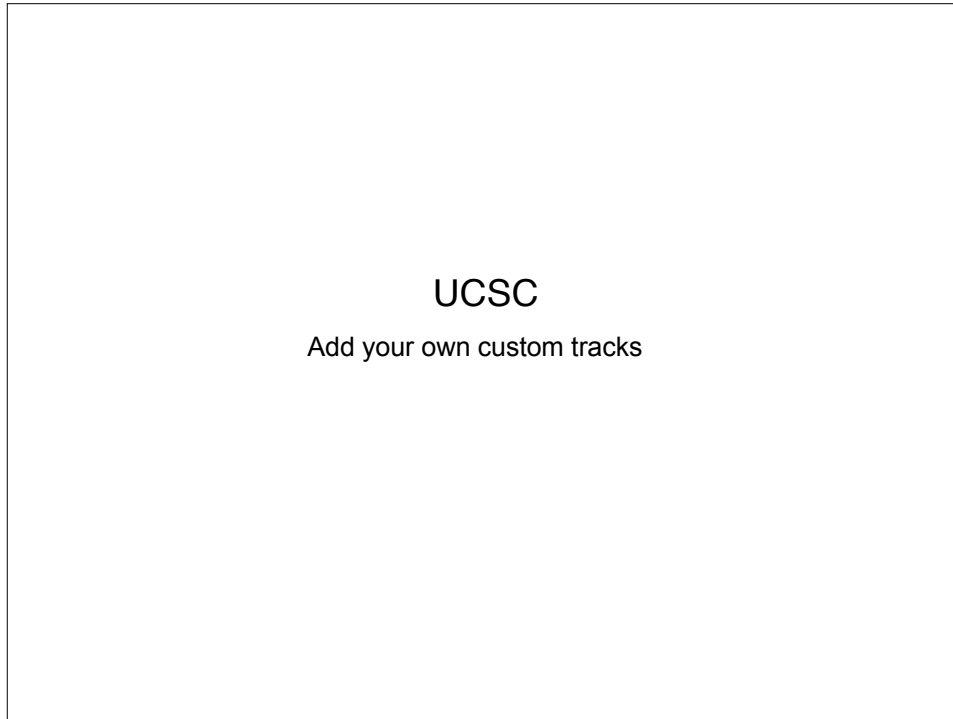
File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.
Upload sequence: Browse... submit file

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.



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, GTF, or BigWig format. 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](#)

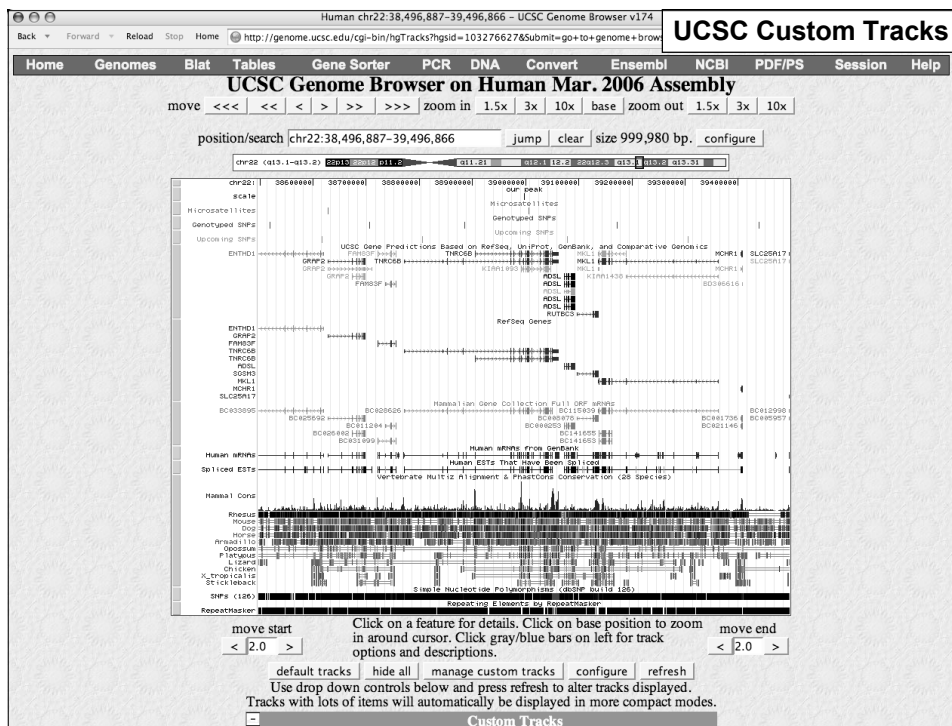
Manage Custom Tracks

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

Name	Description	Type	Doc	Items	Pos	delete	
Upcoming SNPs	Upcoming SNPs	bed		4	chr22:	<input type="checkbox"/>	go to genome browser
Genotyped SNPs	Genotyped SNPs	bed		5	chr22:	<input type="checkbox"/>	go to table browser
Microsatellites	Microsatellites	bed		2	chr22:	<input type="checkbox"/>	
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: genome: assembly:

group: track:

table: [describe table schema](#)

region: genome position [lookup](#) [define regions](#)

identifiers (names/accessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

correlation: [create](#)

output format: Send output to [Gala](#)

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 AND

name does * AND

chrom does * AND

strand does * AND

txStart is AND

txEnd is AND

cdsStart is AND

cdsEnd is AND

exonCount is AND

exonStarts does *

exonEnds does *

id is AND

name2 does * AND

cdsStartStat does * AND

cdsEndStat does * AND

#filter: refGene.exonCount = 1															
bin	name	chrom	strand	txStart	txEnd	cdsStart	cdsEnd	exonCount	exonStarts	exonEnds	id	name2	cdsStartStat	cdsEndStat	exonFrames
585	NM_00100541	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	incompl	0,
587	NM_0010052	chr1	+	357521	358458	357521	358458	1	357521	358458	0	OR4F6	cmpl	incompl	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	incompl	cmpl	0,
589	NM_0010052	chr1	-	610960	611897	610960	611897	1	610960	611897	0	OR4F16	incompl	cmpl	0,
593	NM_080605	chr1	+	1157491	1160281	1157521	1158511	1	1157491	1160281	0	BSGAL16	cmpl	cmpl	0,
607	NM_080431	chr1	+	2927905	2929325	2928110	2929244	1	2927905	2929325	0	ACTR12	cmpl	cmpl	0,
88	NM_006511	chr1	+	15858950	15860803	15858950	15860803	1	15858950	15860803	0	KSC1A1	cmpl	incompl	0,
707	NM_00108951	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	🔍 B G
		Build 35.1	🔍 B
<i>Macaca mulatta</i>	rhesus macaque	Build 1.1	🔍 B G
<i>Pan troglodytes</i>	chimpanzee	Build 2.1	🔍 B G
Rodents (2)			
<i>Mus musculus</i>	laboratory mouse	Build 37.1	🔍 B G
		Build 36.1	🔍 B
<i>Rattus norvegicus</i>	rat	RGSC v3.4	🔍 B G
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	🔍 B G
<i>Aspergillus niger</i>		Build 1.1	🔍 B G
<i>Candida glabrata</i>		Build 1.1	🔍 B
<i>Cryptococcus neoformans</i>		Build 2.1	🔍 B
<i>Debaryomyces hansenii</i>		Build 1.1	🔍 B
<i>Encephalitozoon cuniculi</i>		Build 1.1	🔍 B
<i>Eremothecium gossypii</i>		Build 3.1	🔍 B
<i>Gibberella zeae</i>		Build 1.2	🔍 B
<i>Kluyveromyces lactis</i>		Build 1.1	🔍 B
<i>Magnaporthe oryzae</i>	rice blast fungus	Build 3.1	🔍 B
<i>Neurospora crassa</i>		Build 1.1	🔍 B
<i>Pichia stipitis</i>		Build 1.1	🔍 B
<i>Saccharomyces cerevisiae</i>	baker's yeast	Build 2.1	🔍 B G
<i>Schizosaccharomyces pombe</i>	fission yeast	Build 1.1	🔍 B
<i>Ustilago maydis</i>		Build 1.1	🔍 B
<i>Yarrowia lipolytica</i>		Build 1.1	🔍 B

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

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	D8S1170	STS	STS
8	Celera	all matches	D8S94	D8S1170	STS	STS

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

NCBI NCBI Map Viewer

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]

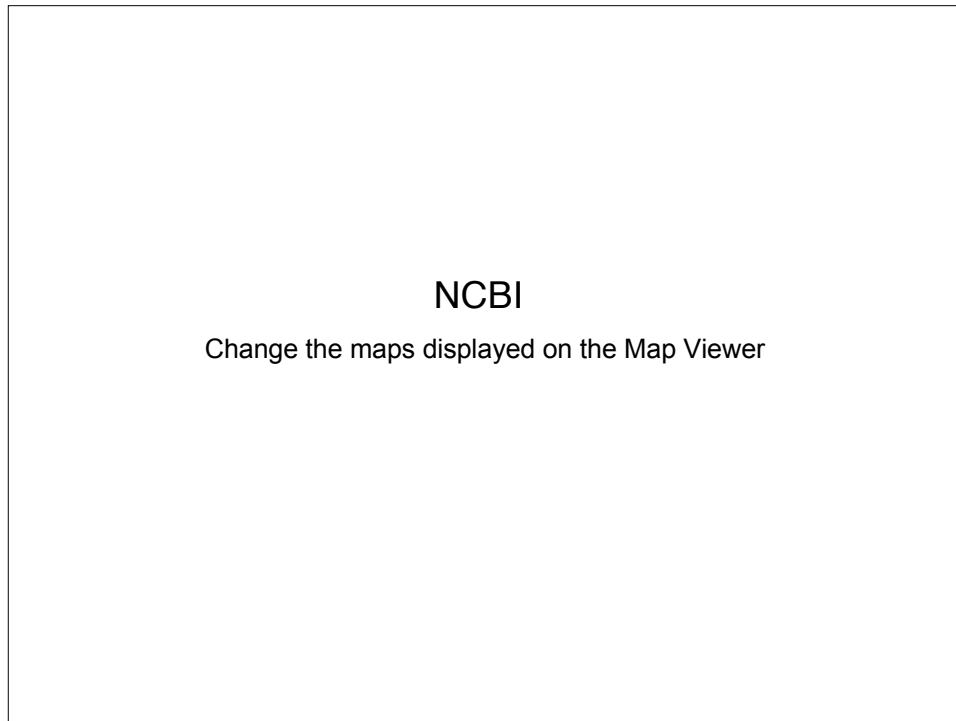
Region Displayed: 32M-44,400K bp

Marker	STS	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
Histogram Data: Tick Width=16,645bp/pixel, Max Height=673 transcripts (logarithmic scale)
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

Back Forward

NCBI PubMed

Search

Human genome overview page (Build 36.2)
Human genome overview page (Build 35.1)
Map Viewer Home

Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options **click**
Compress Map
Region Shown: 32M 44,400K Go

You are here:
Ideogram
8p23
8p22
8p21
8p20
8p11
8p11.1
8p11.2
8p11.3
8p11.4
8p11.5
8p11.6
8p11.7
8p11.8
8p11.9
8p11.10
8p11.11
8p11.12
8p11.13
8p11.14
8p11.15
8p11.16
8p11.17
8p11.18
8p11.19
8p11.20
8p11.21
8p11.22
8p11.23
8p11.24
8p11.25
8p11.26
8p11.27
8p11.28
8p11.29
8p11.30
8p11.31
8p11.32
8p11.33
8p11.34
8p11.35
8p11.36
8p11.37
8p11.38
8p11.39
8p11.40
8p11.41
8p11.42
8p11.43
8p11.44
8p11.45
8p11.46
8p11.47
8p11.48
8p11.49
8p11.50
8p11.51
8p11.52
8p11.53
8p11.54
8p11.55
8p11.56
8p11.57
8p11.58
8p11.59
8p11.60
8p11.61
8p11.62
8p11.63
8p11.64
8p11.65
8p11.66
8p11.67
8p11.68
8p11.69
8p11.70
8p11.71
8p11.72
8p11.73
8p11.74
8p11.75
8p11.76
8p11.77
8p11.78
8p11.79
8p11.80
8p11.81
8p11.82
8p11.83
8p11.84
8p11.85
8p11.86
8p11.87
8p11.88
8p11.89
8p11.90
8p11.91
8p11.92
8p11.93
8p11.94
8p11.95
8p11.96
8p11.97
8p11.98
8p11.99
8p11.100
8p11.101
8p11.102
8p11.103
8p11.104
8p11.105
8p11.106
8p11.107
8p11.108
8p11.109
8p11.110
8p11.111
8p11.112
8p11.113
8p11.114
8p11.115
8p11.116
8p11.117
8p11.118
8p11.119
8p11.120
8p11.121
8p11.122
8p11.123
8p11.124
8p11.125
8p11.126
8p11.127
8p11.128
8p11.129
8p11.130
8p11.131
8p11.132
8p11.133
8p11.134
8p11.135
8p11.136
8p11.137
8p11.138
8p11.139
8p11.140
8p11.141
8p11.142
8p11.143
8p11.144
8p11.145
8p11.146
8p11.147
8p11.148
8p11.149
8p11.150
8p11.151
8p11.152
8p11.153
8p11.154
8p11.155
8p11.156
8p11.157
8p11.158
8p11.159
8p11.160
8p11.161
8p11.162
8p11.163
8p11.164
8p11.165
8p11.166
8p11.167
8p11.168
8p11.169
8p11.170
8p11.171
8p11.172
8p11.173
8p11.174
8p11.175
8p11.176
8p11.177
8p11.178
8p11.179
8p11.180
8p11.181
8p11.182
8p11.183
8p11.184
8p11.185
8p11.186
8p11.187
8p11.188
8p11.189
8p11.190
8p11.191
8p11.192
8p11.193
8p11.194
8p11.195
8p11.196
8p11.197
8p11.198
8p11.199
8p11.200
8p11.201
8p11.202
8p11.203
8p11.204
8p11.205
8p11.206
8p11.207
8p11.208
8p11.209
8p11.210
8p11.211
8p11.212
8p11.213
8p11.214
8p11.215
8p11.216
8p11.217
8p11.218
8p11.219
8p11.220
8p11.221
8p11.222
8p11.223
8p11.224
8p11.225
8p11.226
8p11.227
8p11.228
8p11.229
8p11.230
8p11.231
8p11.232
8p11.233
8p11.234
8p11.235
8p11.236
8p11.237
8p11.238
8p11.239
8p11.240
8p11.241
8p11.242
8p11.243
8p11.244
8p11.245
8p11.246
8p11.247
8p11.248
8p11.249
8p11.250
8p11.251
8p11.252
8p11.253
8p11.254
8p11.255
8p11.256
8p11.257
8p11.258
8p11.259
8p11.260
8p11.261
8p11.262
8p11.263
8p11.264
8p11.265
8p11.266
8p11.267
8p11.268
8p11.269
8p11.270
8p11.271
8p11.272
8p11.273
8p11.274
8p11.275
8p11.276
8p11.277
8p11.278
8p11.279
8p11.280
8p11.281
8p11.282
8p11.283
8p11.284
8p11.285
8p11.286
8p11.287
8p11.288
8p11.289
8p11.290
8p11.291
8p11.292
8p11.293
8p11.294
8p11.295
8p11.296
8p11.297
8p11.298
8p11.299
8p11.300
8p11.301
8p11.302
8p11.303
8p11.304
8p11.305
8p11.306
8p11.307
8p11.308
8p11.309
8p11.310
8p11.311
8p11.312
8p11.313
8p11.314
8p11.315
8p11.316
8p11.317
8p11.318
8p11.319
8p11.320
8p11.321
8p11.322
8p11.323
8p11.324
8p11.325
8p11.326
8p11.327
8p11.328
8p11.329
8p11.330
8p11.331
8p11.332
8p11.333
8p11.334
8p11.335
8p11.336
8p11.337
8p11.338
8p11.339
8p11.340
8p11.341
8p11.342
8p11.343
8p11.344
8p11.345
8p11.346
8p11.347
8p11.348
8p11.349
8p11.350
8p11.351
8p11.352
8p11.353
8p11.354
8p11.355
8p11.356
8p11.357
8p11.358
8p11.359
8p11.360
8p11.361
8p11.362
8p11.363
8p11.364
8p11.365
8p11.366
8p11.367
8p11.368
8p11.369
8p11.370
8p11.371
8p11.372
8p11.373
8p11.374
8p11.375
8p11.376
8p11.377
8p11.378
8p11.379
8p11.380
8p11.381
8p11.382
8p11.383
8p11.384
8p11.385
8p11.386
8p11.387
8p11.388
8p11.389
8p11.390
8p11.391
8p11.392
8p11.393
8p11.394
8p11.395
8p11.396
8p11.397
8p11.398
8p11.399
8p11.400
8p11.401
8p11.402
8p11.403
8p11.404
8p11.405
8p11.406
8p11.407
8p11.408
8p11.409
8p11.410
8p11.411
8p11.412
8p11.413
8p11.414
8p11.415
8p11.416
8p11.417
8p11.418
8p11.419
8p11.420
8p11.421
8p11.422
8p11.423
8p11.424
8p11.425
8p11.426
8p11.427
8p11.428
8p11.429
8p11.430
8p11.431
8p11.432
8p11.433
8p11.434
8p11.435
8p11.436
8p11.437
8p11.438
8p11.439
8p11.440
8p11.441
8p11.442
8p11.443
8p11.444
8p11.445
8p11.446
8p11.447
8p11.448
8p11.449
8p11.450
8p11.451
8p11.452
8p11.453
8p11.454
8p11.455
8p11.456
8p11.457
8p11.458
8p11.459
8p11.460
8p11.461
8p11.462
8p11.463
8p11.464
8p11.465
8p11.466
8p11.467
8p11.468
8p11.469
8p11.470
8p11.471
8p11.472
8p11.473
8p11.474
8p11.475
8p11.476
8p11.477
8p11.478
8p11.479
8p11.480
8p11.481
8p11.482
8p11.483
8p11.484
8p11.485
8p11.486
8p11.487
8p11.488
8p11.489
8p11.490
8p11.491
8p11.492
8p11.493
8p11.494
8p11.495
8p11.496
8p11.497
8p11.498
8p11.499
8p11.500
8p11.501
8p11.502
8p11.503
8p11.504
8p11.505
8p11.506
8p11.507
8p11.508
8p11.509
8p11.510
8p11.511
8p11.512
8p11.513
8p11.514
8p11.515
8p11.516
8p11.517
8p11.518
8p11.519
8p11.520
8p11.521
8p11.522
8p11.523
8p11.524
8p11.525
8p11.526
8p11.527
8p11.528
8p11.529
8p11.530
8p11.531
8p11.532
8p11.533
8p11.534
8p11.535
8p11.536
8p11.537
8p11.538
8p11.539
8p11.540
8p11.541
8p11.542
8p11.543
8p11.544
8p11.545
8p11.546
8p11.547
8p11.548
8p11.549
8p11.550
8p11.551
8p11.552
8p11.553
8p11.554
8p11.555
8p11.556
8p11.557
8p11.558
8p11.559
8p11.560
8p11.561
8p11.562
8p11.563
8p11.564
8p11.565
8p11.566
8p11.567
8p11.568
8p11.569
8p11.570
8p11.571
8p11.572
8p11.573
8p11.574
8p11.575
8p11.576
8p11.577
8p11.578
8p11.579
8p11.580
8p11.581
8p11.582
8p11.583
8p11.584
8p11.585
8p11.586
8p11.587
8p11.588
8p11.589
8p11.590
8p11.591
8p11.592
8p11.593
8p11.594
8p11.595
8p11.596
8p11.597
8p11.598
8p11.599
8p11.600
8p11.601
8p11.602
8p11.603
8p11.604
8p11.605
8p11.606
8p11.607
8p11.608
8p11.609
8p11.610
8p11.611
8p11.612
8p11.613
8p11.614
8p11.615
8p11.616
8p11.617
8p11.618
8p11.619
8p11.620
8p11.621
8p11.622
8p11.623
8p11.624
8p11.625
8p11.626
8p11.627
8p11.628
8p11.629
8p11.630
8p11.631
8p11.632
8p11.633
8p11.634
8p11.635
8p11.636
8p11.637
8p11.638
8p11.639
8p11.640
8p11.641
8p11.642
8p11.643
8p11.644
8p11.645
8p11.646
8p11.647
8p11.648
8p11.649
8p11.650
8p11.651
8p11.652
8p11.653
8p11.654
8p11.655
8p11.656
8p11.657
8p11.658
8p11.659
8p11.660
8p11.661
8p11.662
8p11.663
8p11.664
8p11.665
8p11.666
8p11.667
8p11.668
8p11.669
8p11.670
8p11.671
8p11.672
8p11.673
8p11.674
8p11.675
8p11.676
8p11.677
8p11.678
8p11.679
8p11.680
8p11.681
8p11.682
8p11.683
8p11.684
8p11.685
8p11.686
8p11.687
8p11.688
8p11.689
8p11.690
8p11.691
8p11.692
8p11.693
8p11.694
8p11.695
8p11.696
8p11.697
8p11.698
8p11.699
8p11.700
8p11.701
8p11.702
8p11.703
8p11.704
8p11.705
8p11.706
8p11.707
8p11.708
8p11.709
8p11.710
8p11.711
8p11.712
8p11.713
8p11.714
8p11.715
8p11.716
8p11.717
8p11.718
8p11.719
8p11.720
8p11.721
8p11.722
8p11.723
8p11.724
8p11.725
8p11.726
8p11.727
8p11.728
8p11.729
8p11.730
8p11.731
8p11.732
8p11.733
8p11.734
8p11.735
8p11.736
8p11.737
8p11.738
8p11.739
8p11.740
8p11.741
8p11.742
8p11.743
8p11.744
8p11.745
8p11.746
8p11.747
8p11.748
8p11.749
8p11.750
8p11.751
8p11.752
8p11.753
8p11.754
8p11.755
8p11.756
8p11.757
8p11.758
8p11.759
8p11.760
8p11.761
8p11.762
8p11.763
8p11.764
8p11.765
8p11.766
8p11.767
8p11.768
8p11.769
8p11.770
8p11.771
8p11.772
8p11.773
8p11.774
8p11.775
8p11.776
8p11.777
8p11.778
8p11.779
8p11.780
8p11.781
8p11.782
8p11.783
8p11.784
8p11.785
8p11.786
8p11.787
8p11.788
8p11.789
8p11.790
8p11.791
8p11.792
8p11.793
8p11.794
8p11.795
8p11.796
8p11.797
8p11.798
8p11.799
8p11.800
8p11.801
8p11.802
8p11.803
8p11.804
8p11.805
8p11.806
8p11.807
8p11.808
8p11.809
8p11.810
8p11.811
8p11.812
8p11.813
8p11.814
8p11.815
8p11.816
8p11.817
8p11.818
8p11.819
8p11.820
8p11.821
8p11.822
8p11.823
8p11.824
8p11.825
8p11.826
8p11.827
8p11.828
8p11.829
8p11.830
8p11.831
8p11.832
8p11.833
8p11.834
8p11.835
8p11.836
8p11.837
8p11.838
8p11.839
8p11.840
8p11.841
8p11.842
8p11.843
8p11.844
8p11.845
8p11.846
8p11.847
8p11.848
8p11.849
8p11.850
8p11.851
8p11.852
8p11.853
8p11.854
8p11.855
8p11.856
8p11.857
8p11.858
8p11.859
8p11.860
8p11.861
8p11.862
8p11.863
8p11.864
8p11.865
8p11.866
8p11.867
8p11.868
8p11.869
8p11.870
8p11.871
8p11.872
8p11.873
8p11.874
8p11.875
8p11.876
8p11.877
8p11.878
8p11.879
8p11.880
8p11.881
8p11.882
8p11.883
8p11.884
8p11.885
8p11.886
8p11.887
8p11.888
8p11.889
8p11.890
8p11.891
8p11.892
8p11.893
8p11.894
8p11.895
8p11.896
8p11.897
8p11.898
8p11.899
8p11.900
8p11.901
8p11.902
8p11.903
8p11.904
8p11.905
8p11.906
8p11.907
8p11.908
8p11.909
8p11.910
8p11.911
8p11.912
8p11.913
8p11.914
8p11.915
8p11.916
8p11.917
8p11.918
8p11.919
8p11.920
8p11.921
8p11.922
8p11.923
8p11.924
8p11.925
8p11.926
8p11.927
8p11.928
8p11.929
8p11.930
8p11.931
8p11.932
8p11.933
8p11.934
8p11.935
8p11.936
8p11.937
8p11.938
8p11.939
8p11.940
8p11.941
8p11.942
8p11.943
8p11.944
8p11.945
8p11.946
8p11.947
8p11.948
8p11.949
8p11.950
8p11.951
8p11.952
8p11.953
8p11.954
8p11.955
8p11.956
8p11.957
8p11.958
8p11.959
8p11.960
8p11.961
8p11.962
8p11.963
8p11.964
8p11.965
8p11.966
8p11.967
8p11.968
8p11.969
8p11.970
8p11.971
8p11.972
8p11.973
8p11.974
8p11.975
8p11.976
8p11.977
8p11.978
8p11.979
8p11.980
8p11.981
8p11.982
8p11.983
8p11.984
8p11.985
8p11.986
8p11.987
8p11.988
8p11.989
8p11.990
8p11.991
8p11.992
8p11.993
8p11.994
8p11.995
8p11.996
8p11.997
8p11.998
8p11.999
8p11.1000

Org: human Assembly: reference Change Assembly

Available Maps: Phenotype STS Gene NCI Clone Phenotype RefSeq Transcripts Repeats rnaBt rnaGaa

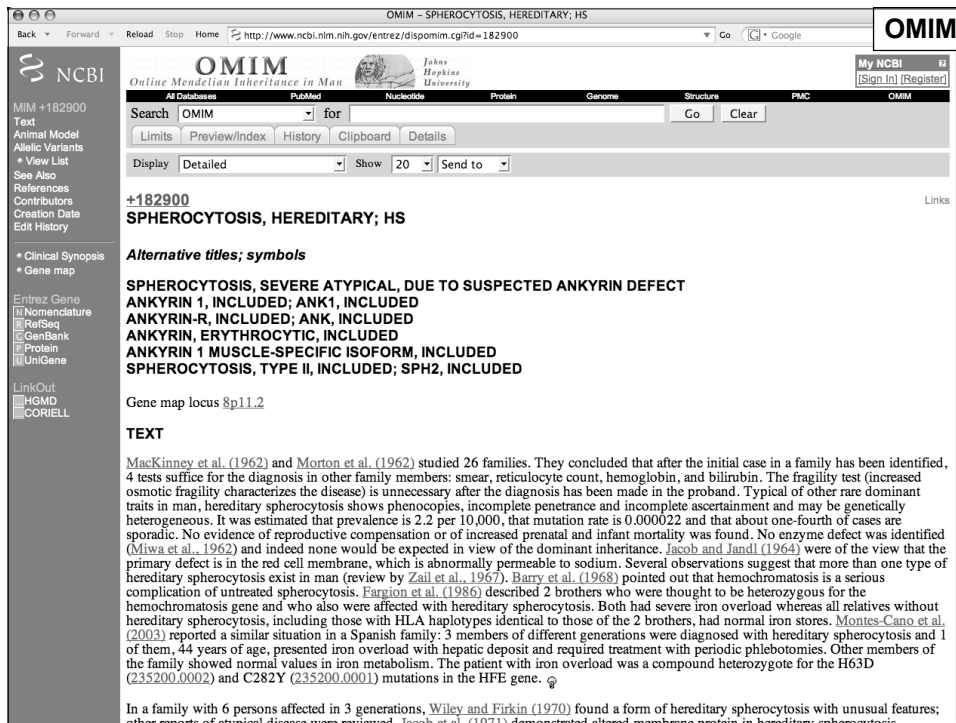
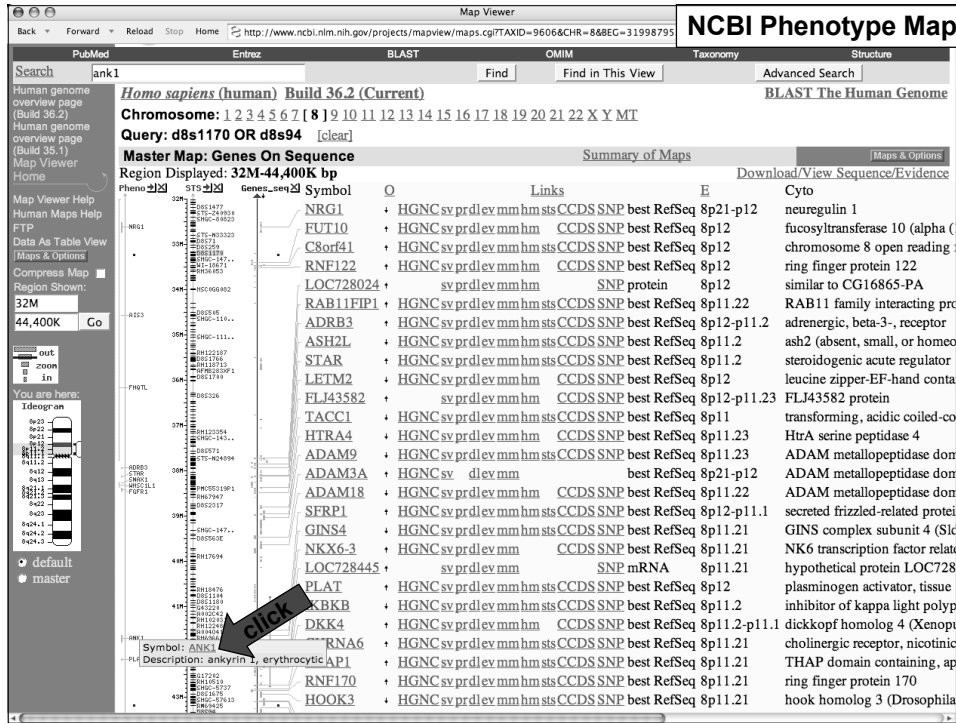
Maps Displayed (left to right): Phenotype STS Gene

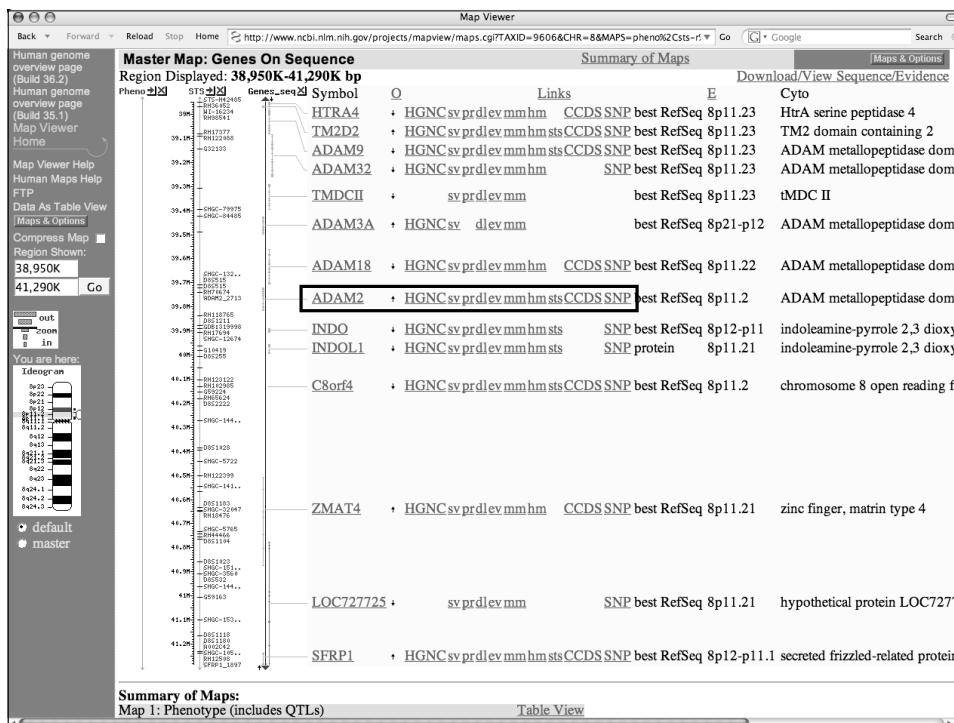
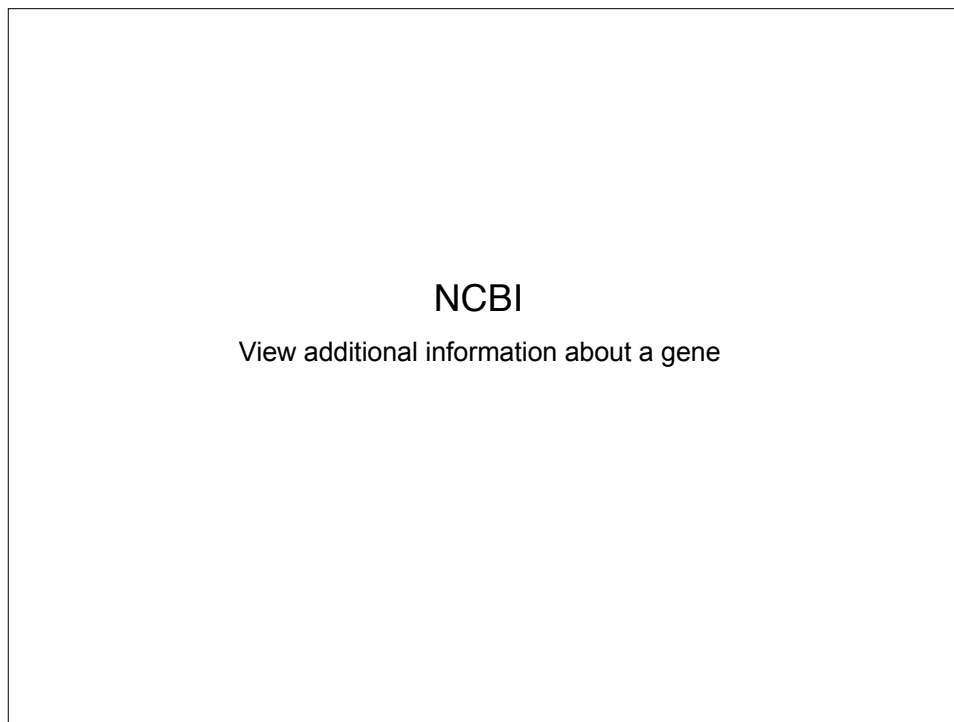
ADD>> <<REMOVE

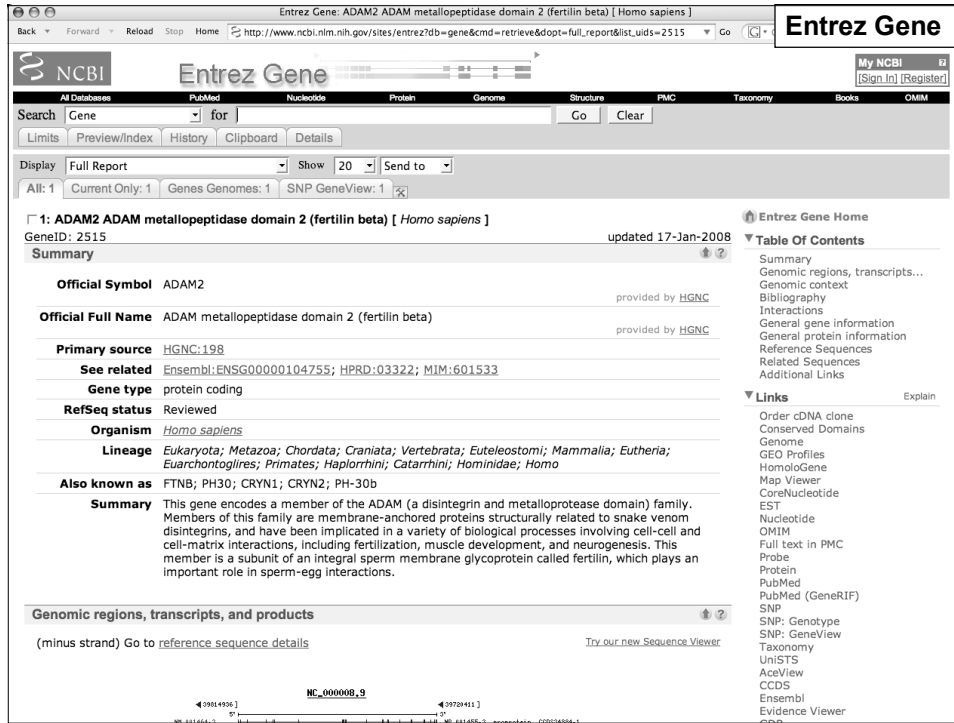
Move UP Move DOWN Make Master/Move to Bottom Toggle Ruler ([R] before map means 'ruler set')

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

OK Apply Close







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*

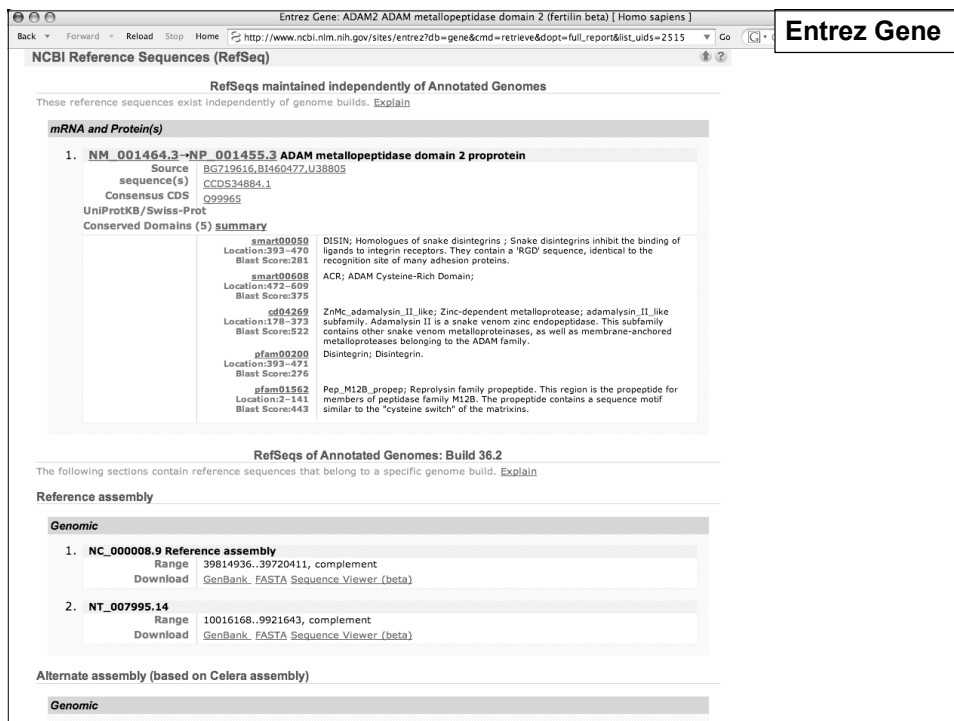
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

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](#)



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 propeptide**

Source sequence(s) BG719616.B1460477.U38805
Consensus CDS CCD534884.1
Consensus CDS Q39965
UniProtKB/Swiss-Prot

Conserved Domains (5) summary

smart00850	Location:393-470 Blast Score:281	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.
smart06008	Location:472-609 Blast Score:375	ACR; ADAM Cysteine-Rich Domain;
cd04269	Location:178-373 Blast Score:322	ZnMc_adamalyisin_II_like; Zinc-dependent metalloprotease; adamalyisin_II_like subfamily. Adamalyisin 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.
pfam09200	Location:393-471 Blast Score:276	Disintegrin; Disintegrin.
pfam01562	Location:2-141 Blast Score:443	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

HomoloGene (hm)

Search: HomoloGene for [] Go Clear

Display: HomoloGene Show 20 Send to []

All: 1 Fungi: 0 Mammals: 1

1: HomoloGene:1127. Gene conserved in Eutheria Download, Links

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

- ADAM2, *Homo sapiens*
ADAM metallopeptidase domain 2 (fertilin beta)
- ADAM2, *Pan troglodytes*
ADAM metallopeptidase domain 2 (fertilin beta)
- ADAM2, *Canis lupus familiaris*
ADAM metallopeptidase domain 2 (fertilin beta)
- Adam2, *Mus musculus*
a disintegrin and metalloproteinase domain 2
- Adam2, *Rattus norvegicus*
a disintegrin and metalloprotease domain 2

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

- NP_001455.3
735 aa
- XP_519722.2
735 aa
- XP_532795.2
881 aa
- NP_033748.2
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.

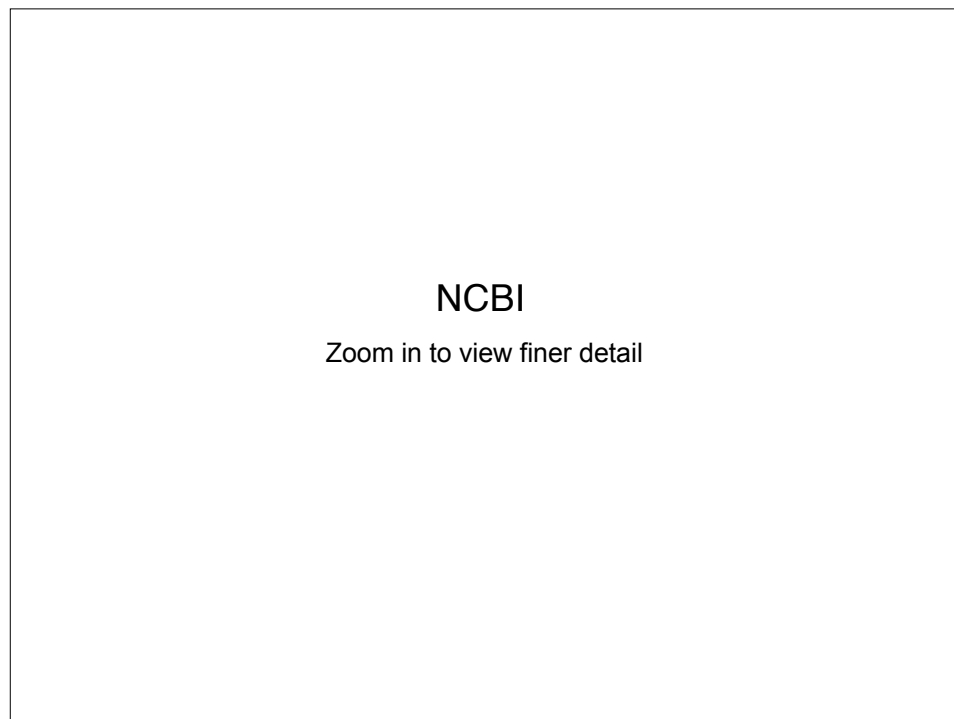
Regenerate Alignments

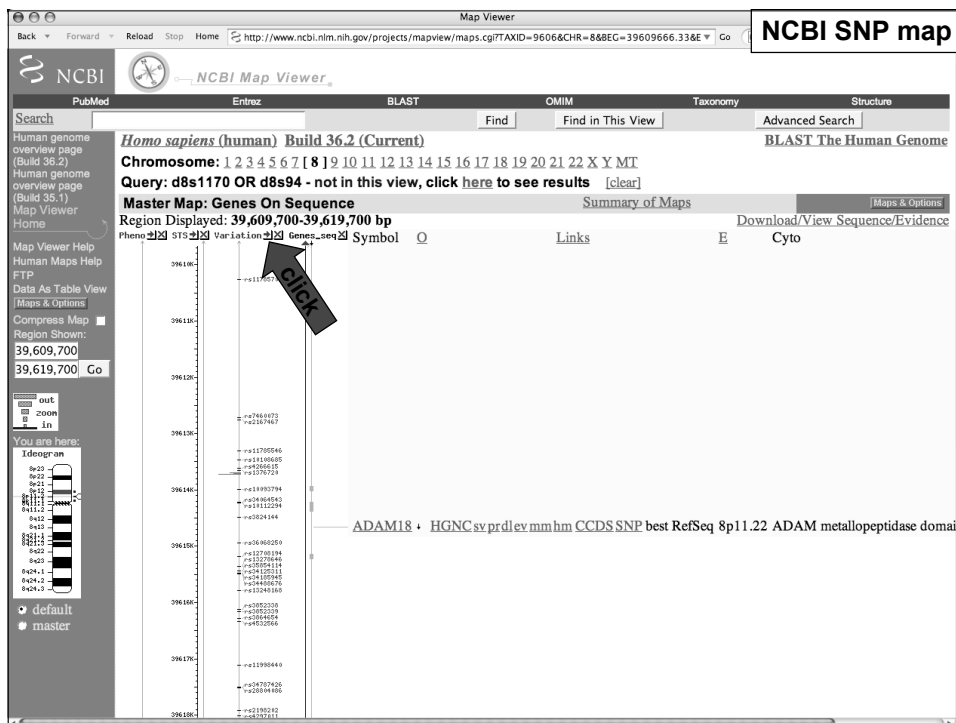
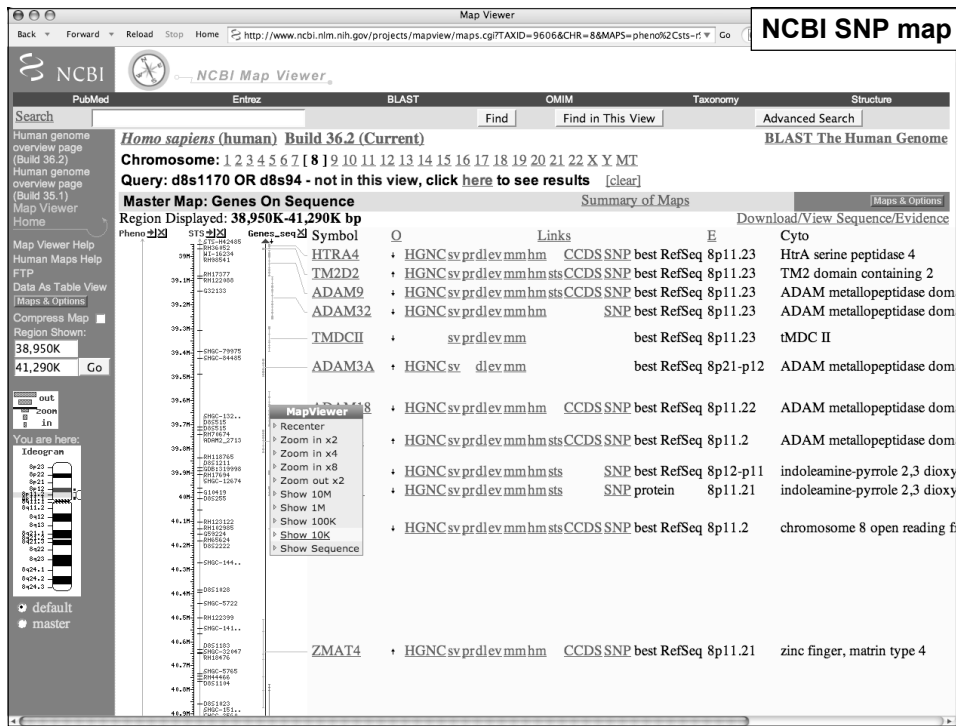
NP_001455.3 (Homo sapiens. ADAM2) [v]
XP_519722.2 (Pan troglodytes. ADAM2) [v]

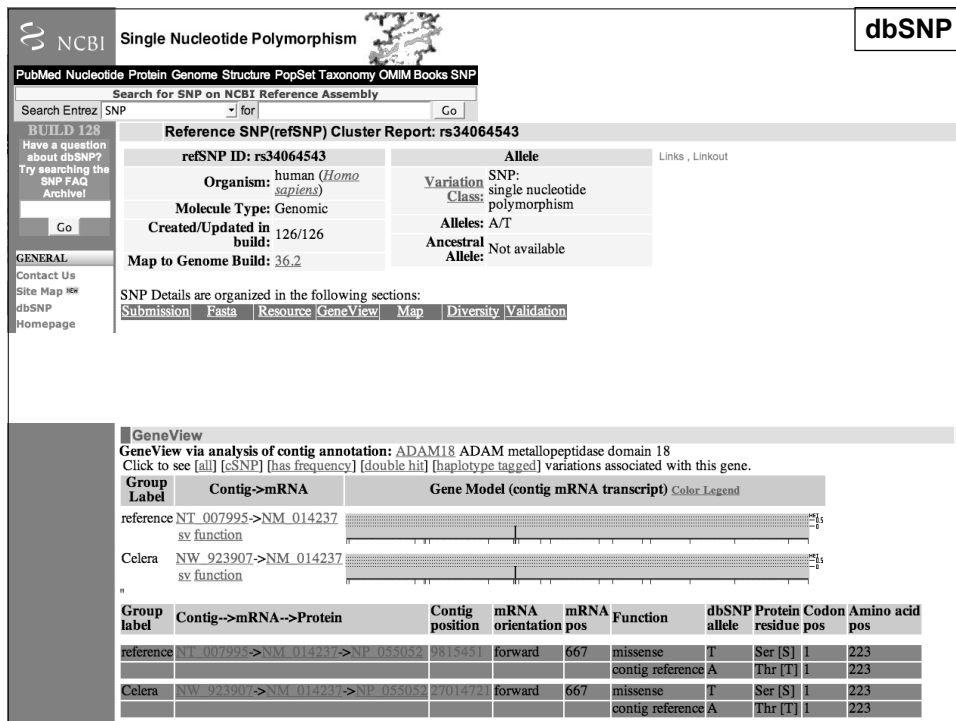
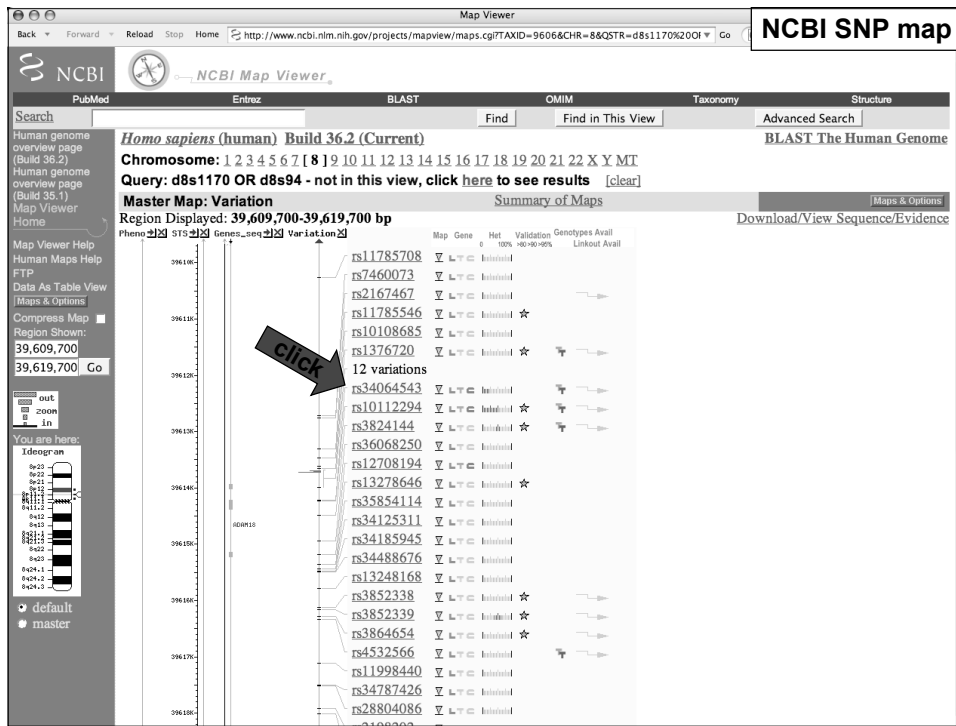
BLAST

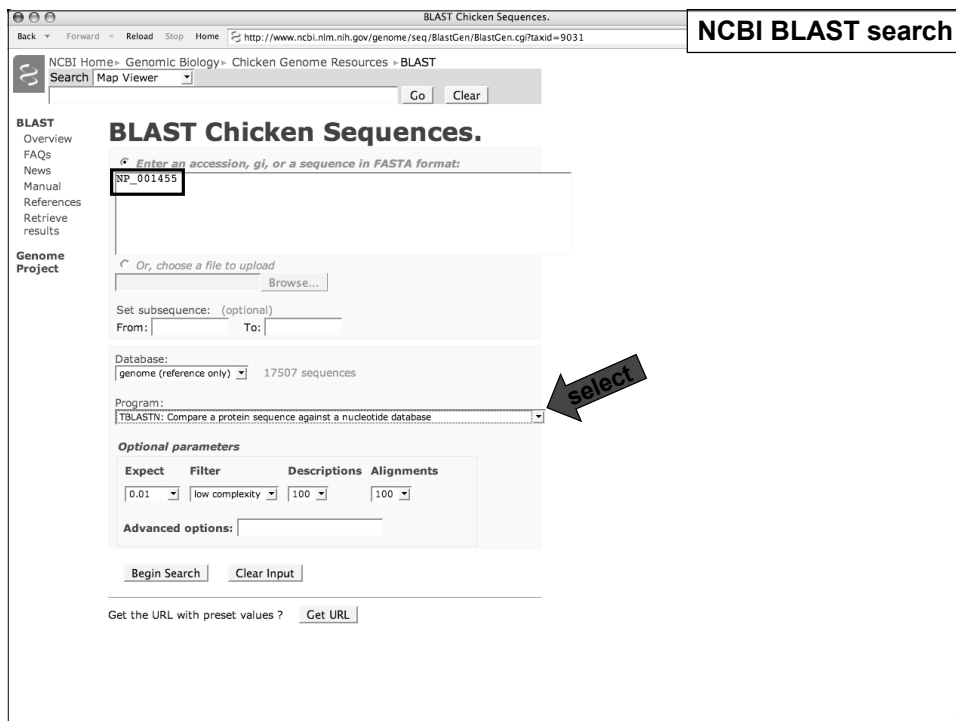
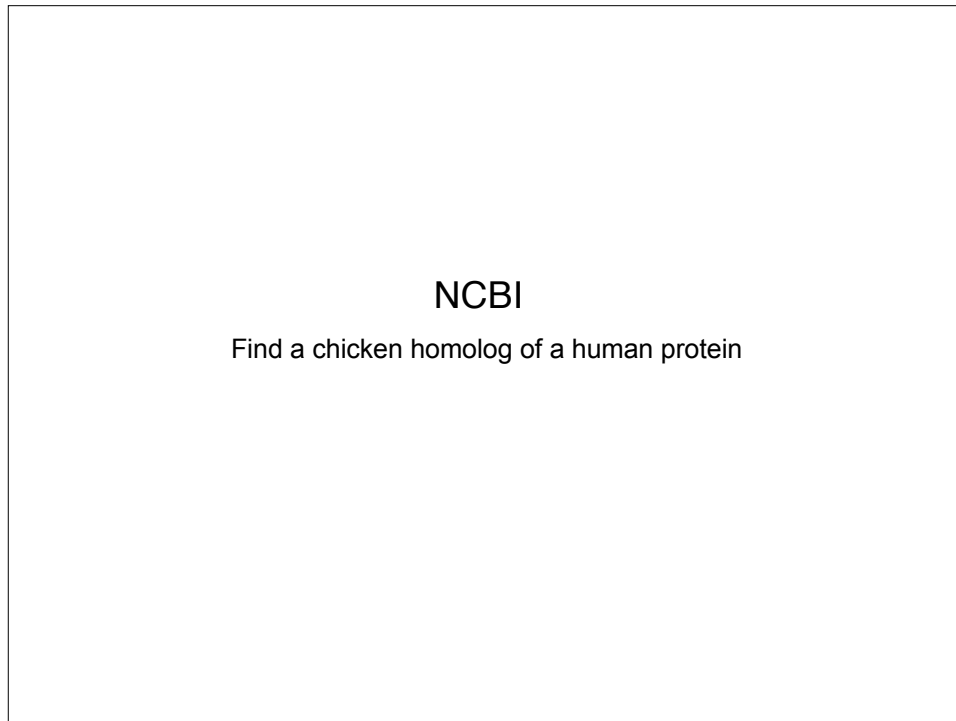
Conserved Domains
Conserved Domains from CDD found in protein sequences by rpsblast searching.

- ZnMe_adamalyisin_IL_IIke (cd04269)
= Zinc-dependent metalloprotease; adamalyisin_IL_IIke subfamily. Adamalyisin II is a snake venom zinc endopeptidase. This subfamily contains other snake venom metalloproteinases, as well as membrane-anchored metalloproteases belonging to the ADAM family.
- Pep_M12B_propep (pfam01562)
= 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.
- ACR (smart00608)









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 proprotein (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|CA049812.1| 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:

ref NW_001471710.1 Gga5_WGA136_2	Gallus gallus chromosome 5 g...	Score	E
ref NW_001471459.1 Gga15_WGA207_2	Gallus gallus chromosome 15...	359	1e-97
ref NW_001471578.1 Gga22_WGA268_2	Gallus gallus chromosome 22...	273	9e-72
		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	ASCPENHYVOTCHPCOLNOMICTDVCMSQDKQCTDFFQKEVEFCPCSECVSHLSKTDVS	522
Sbjct 214257	C E+ YVQ G PC + C G C S KQC PG++ P BC+ +N++ D	
Query 523	GNGGIEDS-GVTQCEADNLQCGKLCIKYVGFLLQIPRATIIYANISGLCAVEFASDH	581
Sbjct 214434	GNGG ++ +T+C +N CG+L C+ V + L 2+I G C +++	
Query 582	ADSQKNWIKDQTSQSNKVCNRQRCVSSVLYDCTCKCNRQVGNKKNKHCBSYLP	641
Sbjct 214614	S ++DGT+CGSNK+C + C + S L YDC KC+RQVGNH R+CIC + P	

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 proprotein (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|CA049812.1| 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
-----	-------	-------	--------	-------

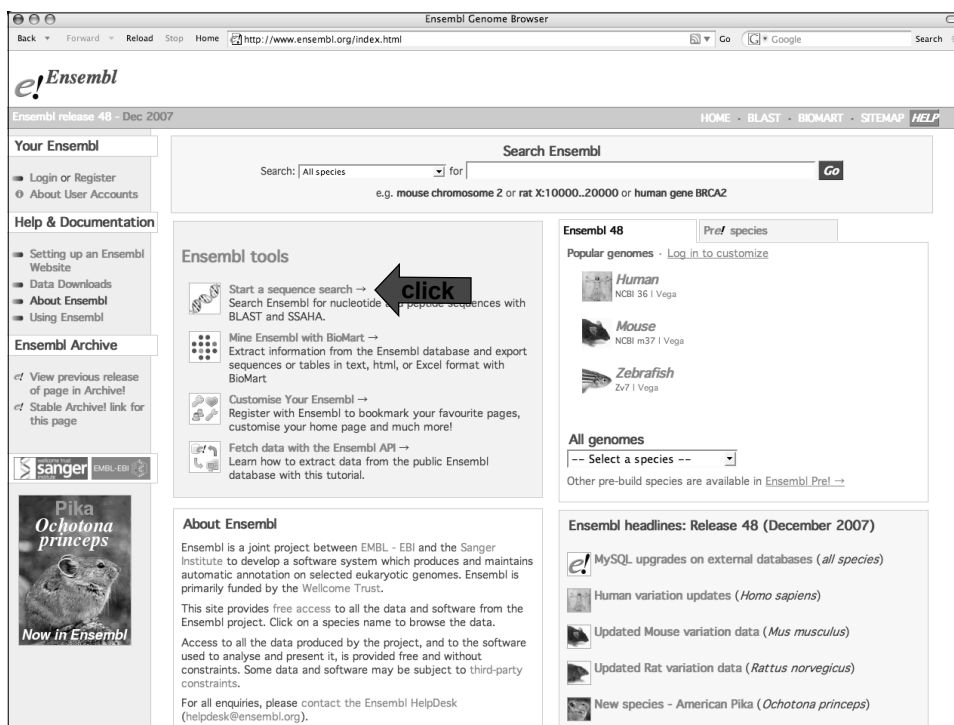
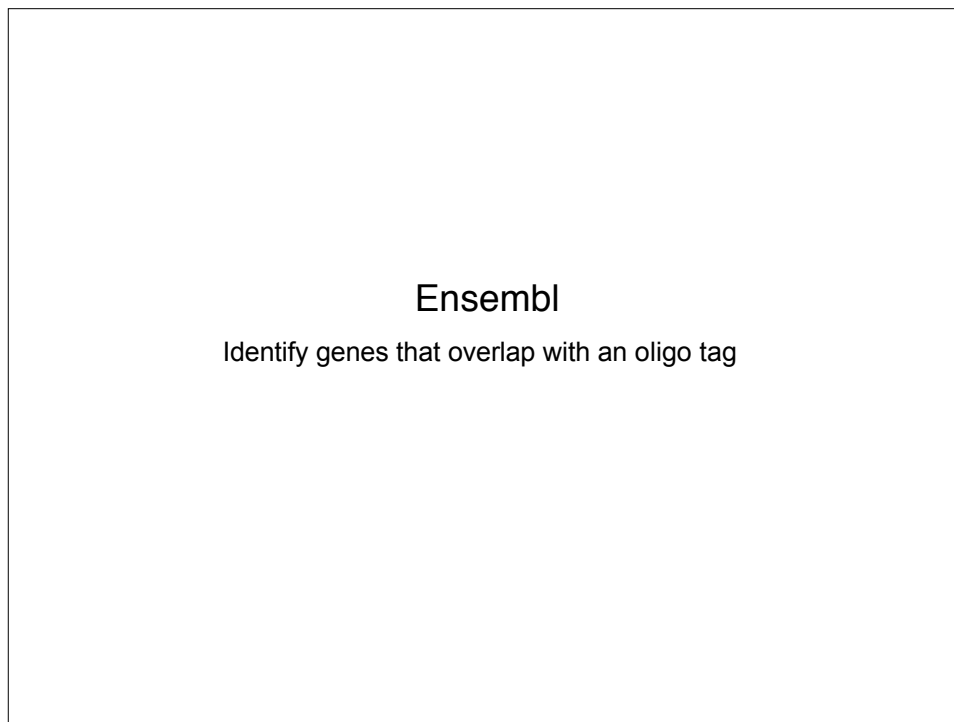
Sequences producing significant alignments:

ref NW_001471710.1 Gga5_WGA136_2	Gallus gallus chromosome 5 g...	Score	E
ref NW_001471459.1 Gga15_WGA207_2	Gallus gallus chromosome 15...	359	1e-97
ref NW_001471578.1 Gga22_WGA268_2	Gallus gallus chromosome 22...	273	9e-72
		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	ASCPENHYVOTCHPCOLNOMICTDVCMSQDKQCTDFFQKEVEFCPCSECVSHLSKTDVS	522
Sbjct 214257	C E+ YVQ G PC + C G C S KQC PG++ P BC+ +N++ D	
Query 523	GNGGIEDS-GVTQCEADNLQCGKLCIKYVGFLLQIPRATIIYANISGLCAVEFASDH	581
Sbjct 214434	GNGG ++ +T+C +N CG+L C+ V + L 2+I G C +++	
Query 582	ADSQKNWIKDQTSQSNKVCNRQRCVSSVLYDCTCKCNRQVGNKKNKHCBSYLP	641
Sbjct 214614	S ++DGT+CGSNK+C + C + S L YDC KC+RQVGNH R+CIC + P	

34



Ensembl BLAST search

Ensembl release 48 - Dec 2007

Your Ensembl

- Login or Register
- About User Accounts

Enter the Query Sequence

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

```
MPSS_1
AAAAATGTCCGCCTGAAGAG
```

Or Upload a file containing one or more FASTA sequences

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

Or Enter an existing ticket ID:

Select the databases to search against

Select species: Gallus_gallus, Gasterosteus_aculeatus, Homo_sapiens

Select the Search Tool: BLASTN, BLAT, SSAHA2

Search sensitivity: Near-exact matches (oligo), Exact matches, Near-exact matches, Near-exact matches (oligo), Allow some local mismatch, Distant homologies

Summary

- setup (Not yet initialised)
- configure (Not yet initialised)
- results (Not yet initialised)
- display (Not yet initialised)

Ensembl BLAST search

100% identity over 100% of the query length

Query Start: 1, End: 20, % ID: 100.00

Alignment Summary (click arrow to hide)

Links	Query	Chromosome	Name	Start	End	Ori	Chromosome	Name	Start	End	Ori	Score	E-val	PIID	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

Ensembl ContigView
Ensembl release 48 - Dec 2007

Chromosome 15

Overview

Chr 15 band: p13, p12, q11.2, q14, q21.1, q23, q25.1

Chr 15 band: 54.50 Mb, 54.70 Mb, 54.90 Mb, 55.10 Mb, 55.30 Mb

DNA(cortigs): D15SS24, D15S1225, AFM021a5, D15S848, D15S1127, D15S1169

Markers: D15SS24, D15S1225, AFM021a5, D15S848, D15S1127, D15S1169

Ensembl Genes: TNSI, SUH4W, NOVEL, TCF12, UGN

Gene legend: Ensembl Known Protein Coding, Unassigned Known Protein coding, RNA Pseudogene (Novel), Ensembl Novel intron retained

Detailed view

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

Chr 15 Length: 54,996,500 - 54,997,500 - 54,998,500 - 54,999,500

Conservation: Conserved elements

EST trans: ENSEST000000031.720 >

Ensembl trans: TCF12 >

EST hits: NP_366520.1 >

DNA(cortigs): AC010668.1:22,186 >

Human tlepath clones: FE11-323F24 >

Reg. Features: Length: 54,996,500 - 54,997,500 - 54,998,500 - 54,999,500

Gene legend: Common Known Protein coding, EST gene, Unclassified, Promoter associated

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

Basepair view

Chr 15 Length: 54,996,130 - 54,996,170 - 54,996,190 - 54,996,210

Conservation: Conserved elements

EST trans: ENSEST000000031.720 >

Ensembl trans: TCF12 >

EST hits: NP_366520.1 >

DNA(cortigs): AC010668.1:22,186 >

Human tlepath clones: FE11-323F24 >

Reg. Features: Length: 54,996,130 - 54,996,170 - 54,996,190 - 54,996,210

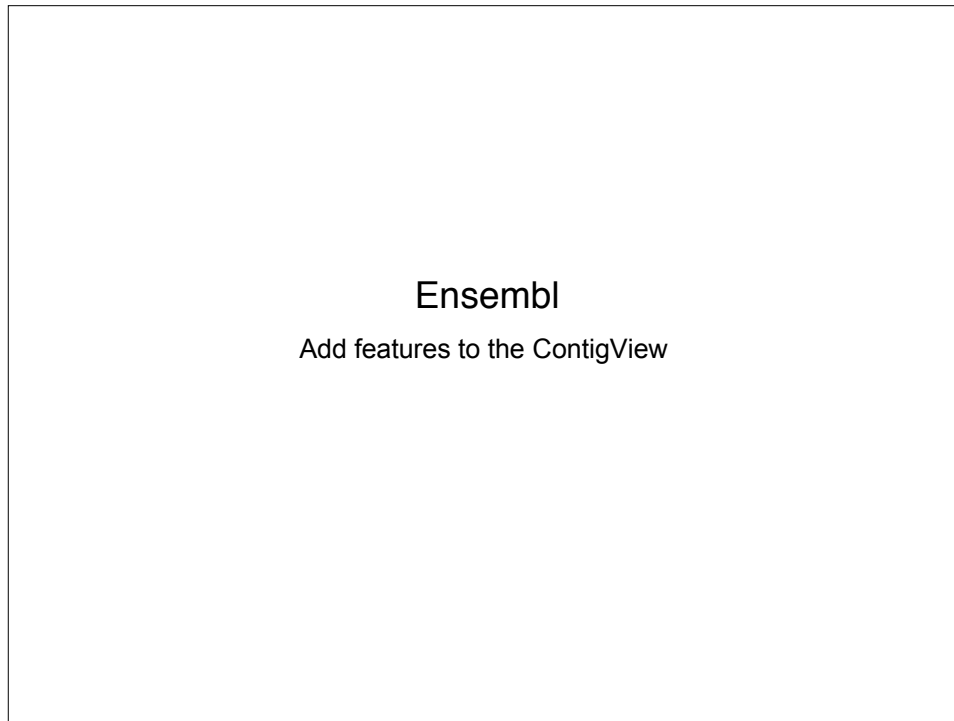
Gene legend: Common Known Protein coding, EST gene, Unclassified, Promoter associated

Reg. feats legend: Promoter associated

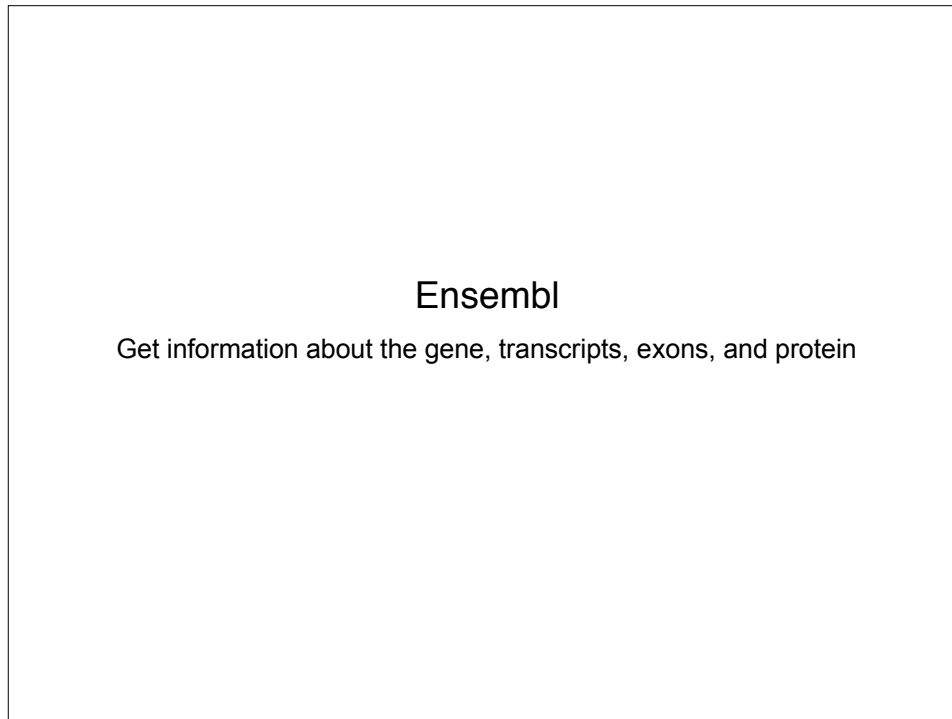
These are currently 126 tracks switched off, use the menu above the image to turn them on.
Ensembl Homo sapiens version 48.36 (NCBI 36) Chromosome 15 54,996,168 - 55,000,187

Ensembl Homo sapiens version 48.36 (NCBI 36) Chromosome 15 54,996,128 - 54,998,227

© 2008 WTSI / EBI. Ensembl is available to download for public use - please see the code licence for details.



The screenshot displays the Ensembl ContigView interface for Chromosome 15, region 54,996,168 - 55,000,187. The interface includes a navigation menu on the left with options like 'Your Ensembl', 'Chromosome 15', 'Overview', and 'Detailed view'. A 'select' arrow points to the 'Detailed view' option. The main content area shows a genomic track with various features such as SNPs, CpG islands, and genes. A legend on the right lists features like 'Common Known Protein coding', 'EST gene', and 'Promoter associated'. The interface also includes search bars, zoom controls, and a 'Refresh' button.



The screenshot displays the Ensembl ContigView interface for Chromosome 15. The browser address bar shows the URL: http://www.ensembl.org/Homo_sapiens/contigview?panel_top=off;... The page title is "Ensembl Human ContigView". The left sidebar contains navigation options like "Your Ensembl", "Chromosome 15", and "Overview". The main content area shows a genomic track with various annotations including "Forward strand", "Conserved elements", "EST trans.", "Ensembl trans.", "Blast hits", "DNA (contigs)", "Human Mespach clones", "SNPs", and "Reg. Features". A "click" arrow points to a specific gene feature. The bottom of the page includes a copyright notice: "© 2008 WTSI / EBI. Ensembl is available to download for public use - please see the page licence for details."

The screenshot shows the Ensembl GeneView page for the gene TCF12 (ENSG00000140262). The page includes a navigation sidebar on the left with options like 'Gene information', 'Gene regulation info', and 'Genomic sequence'. The main content area displays the gene's details, including its location on Chromosome 15 (54,998,125 - 55,368,004) and its description as a transcription factor. A 'Transcripts' table lists several transcripts with their IDs and coordinates. Below this, a 'Features' section shows a genomic track with exons and introns for various transcripts. An 'Alignments' section lists other species for which genomic alignments are available, such as Rattus norvegicus and Mus musculus. A black arrow with the word 'click' points to the 'Rattus norvegicus' link in the 'Alignments' list.

The screenshot shows the Ensembl AlignSliceView interface for Chromosome 15. The top navigation bar includes 'Overview' and 'Detailed view' tabs. The 'Detailed view' tab is active, showing a genomic track for the region 54,998,125 to 55,368,004. The track displays various genomic features, including transcripts (e.g., ENSEST0000001725, ENSEST0000001727), protein-coding genes (e.g., TCF12), and other annotations. A 'Features' dropdown menu is visible at the top of the track, and a 'Zoom' control is present. The bottom of the page includes a 'Basepair view' section and a copyright notice for 2008 WTSI / EBI.

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: RId: 29; Query: RId: 22]
<i>Anopheles gambiae</i>	1-to-many	AGAP008814 (Novel Ensembl prediction) [FullContextView] [Align] No description [Target: RId: 31; Query: RId: 22]
<i>Bos taurus</i>	1-to-1	ENSBTAG0000000586 (AJINB4_BOVIN) [FullContextView] [Align] HSC142325 protein. [Source: Uniprot/SPTREMBL; Acc: A0JN64] [Target: RId: 97; Query: RId: 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 co-expressed 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-Der-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 PVD/HSN/PHE 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 (VU) 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 VU precursors during the lateral specification event that distinguishes these two cell fates. [Source: WormBase] [Target: RId: 24; Query: RId: 13]
<i>Canis familiaris</i>	1-to-1	ENSCAF00000016200 (TCF12) [FullContextView] [Align] Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SwissProt; Acc: Q99081] [Target: RId: 96; Query: RId: 82]
<i>Cavia porcellus</i>	1-to-1	ENSFCPG00000009153 (TCF12) [FullContextView] [Align] Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SwissProt; Acc: Q99081] [Target: RId: 82; Query: RId: 79]
<i>Ciona intestinalis</i>	1-to-many	ENSING00000009523 (Q4HN7_CIOIN) [FullContextView] [Align] Transcription factor protein. [Source: Uniprot/SPTREMBL; Acc: Q4HN7] [Target: RId: 24; Query: RId: 21]
<i>Ciona savignyi</i>	1-to-many	ENSICAVG00000011705 (Novel Ensembl prediction) [FullContextView] [Align] No description [Target: RId: 22; Query: RId: 20]
<i>Danio rerio</i>	1-to-1	ENSDBG00000004714 (zgc:85956) [FullContextView] [Align] transcription factor 12 [Source: RefSeq; Accession: NP_299281.1] [Target: RId: 64; Query: RId: 63]
<i>Dasypus novemcinctus</i>	1-to-1	ENSNDG00000003864 (TCF12) [FullContextView] [Align] Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SwissProt; Acc: Q99081] [Target: RId: 59; Query: RId: 58]
<i>Drosophila melanogaster</i>	1-to-many	CG5112 (db) [FullContextView] [Align] Fushi4 daughterless. [Source: Uniprot/SwissProt; Acc: P11420] [Target: RId: 24; Query: RId: 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/SwissProt; Acc: Q99081] [Target: RId: 72; Query: RId: 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/SwissProt; Acc: Q99081] [Target: RId: 70; Query: RId: 70]
<i>Felis catus</i>	1-to-1	ENSFCAG00000001867 (TCF12) [FullContextView] [Align] Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). [Source: Uniprot/SwissProt; Acc: Q99081] [Target: RId: 93; Query: RId: 60]

Ensembl release 48: Homo sapiens Gene report for ENSG00000140262

Ensembl GeneView

Ensembl Human GeneView

Search: e.g. ENSG00000139618, ENSG00000138411

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 info
- ID history
- Compare SNPs in transcript
- Transcript information
- 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 [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; Acc: Q99081]

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

Transcripts

Transcript ID	Transcript Name	Transcript Length	Transcript Info	Exon Info	Peptide Info
TCF12	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 Oryzctolagus 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 Ornithorhynchus anatinus](#)
- [view genomic alignment with Bos taurus](#)

Ensembl release 48: Homo sapiens Gene report for ENSG00000140262

Ensembl GeneView

Search: >

e.g. ENSG00000139618, ENSG00000138411

HOME - BLAST - BIOMART - SITEMAP - HELP

Ensembl release 48 - Dec 2007

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 information
- 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 AC010999.6.1.221986**.

Description: Transcription factor 12 (Transcription factor HTF-4) (E-box-binding protein) (DNA-binding protein HTF4). *Source: Uniprot/ENSRIB/Spot 29901*

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	ENST0000033725	ENSP00000331057	[Transcript info]	[Exon info]	[Peptide info]
NP_996923.1	ENST0000034827	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 domestica
- view genomic alignment with Leontidea africana
- view genomic alignment with Echinos telfairi
- view genomic alignment with Oryctolagus cuniculus
- view genomic alignment with Dasyatis rostriductus
- view genomic alignment with Canis familiaris
- view genomic alignment with Pan troglodytes
- view genomic alignment with Gallus gallus
- view genomic alignment with Oryzohyrinchus anathus
- view genomic alignment with Bos taurus

Ensembl release 48: Homo sapiens Exon Report for ENST00000267811

Ensembl ExonView

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

aaatagtagttttacattctctcttcaaatataatgattttttg.....

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: 100 -99 -97 -95 -90 -85 -80 NO EVIDENCE

NM_207026.1 gH457007[Ensembl_207026.1] Homo sapiens transcription factor 12 (HTF4, heix-loop-helix transcription factor 4) (TCF12), transcript variant 1, mRNA

NM_002025.3 gH457007[Ensembl_002025.3] Homo sapiens transcription factor 12 (HTF4, heix-loop-helix transcription factor 4) (TCF12), transcript variant 3, mRNA

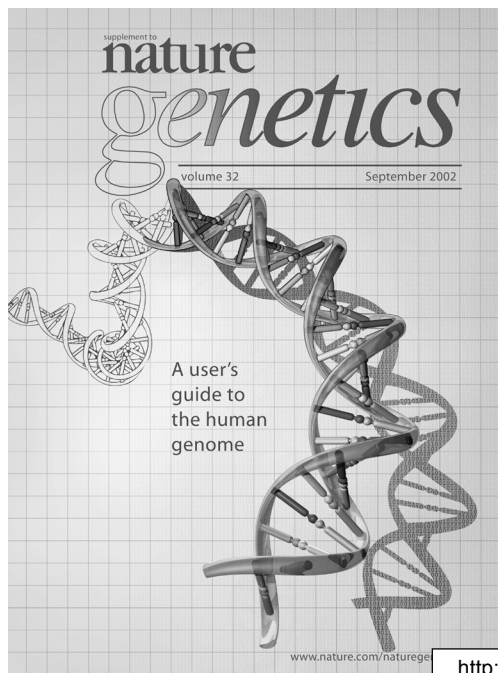
BC051769.2 BC051769.2 Homo sapiens cDNA clone IMAGE:575203, containing frame-shift errors

BC050596.1 BC050596.1 Homo sapiens transcription factor 12 (HTF4, heix-loop-helix transcription factor 4), mRNA in DNA clone IMC-5760 (IMAGE:576757), complete cds

BC13097.1

Additional resources

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



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

Current Protocols in Bioinformatics	
<p>The UCSC Genome Browser</p> <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 (Hinrichs et al., 2003) and a 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 and its annotations.</p>	<p>UNIT 1.4</p>
<p>Using the NCBI Map Viewer to Browse Genomic Sequence Data</p> <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 genome 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 and its annotations. It also describes some of NCBI's other tools, which are provided as links from the Map Viewer. The Alternate Protocols 1 and 2 show how to query the genome sequence, and also illustrate how to perform a BLAST search against the human genome. Alternate Protocol 3 shows how to find all annotated members of a gene.</p> <p>NCBI provides Map Viewers for eleven vertebrates, six invertebrates, 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>	<p>UNIT 1.5</p>
<p>Using the Ensembl Genome Server to Browse Genomic Sequence Data</p> <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>	<p>UNIT 1.15</p>
<p>Access through http://nihlibrary.nih.gov/ResearchTools/OnlineJournals.htm</p>	

Current Topics in Genome Analysis

Next Lecture:

Evolutionary Analysis

Fiona Brinkman, Ph.D.
Simon Fraser University