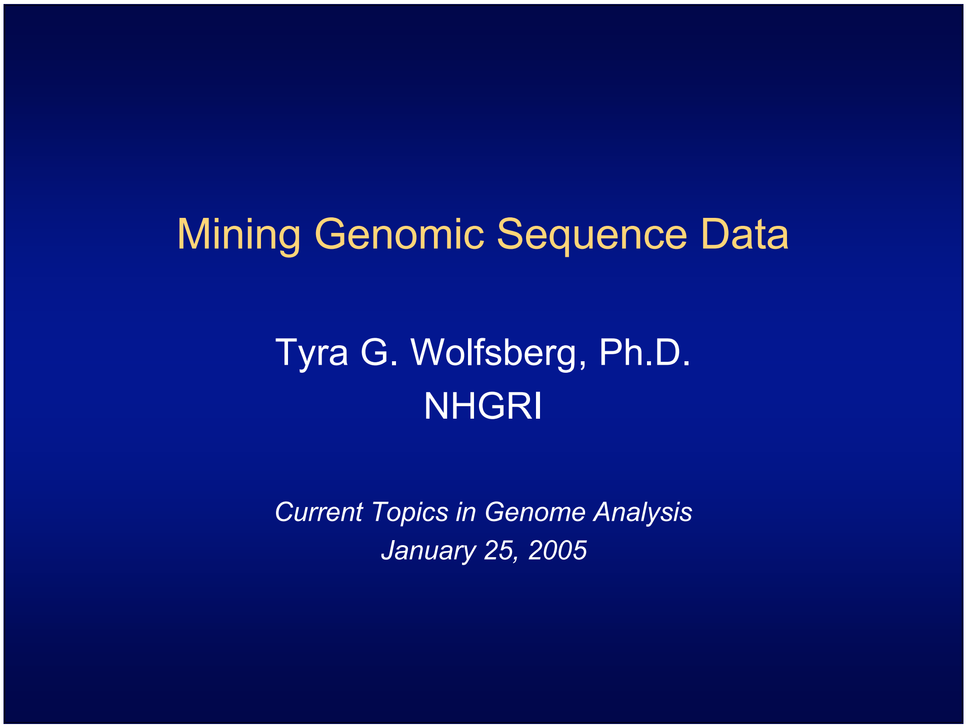
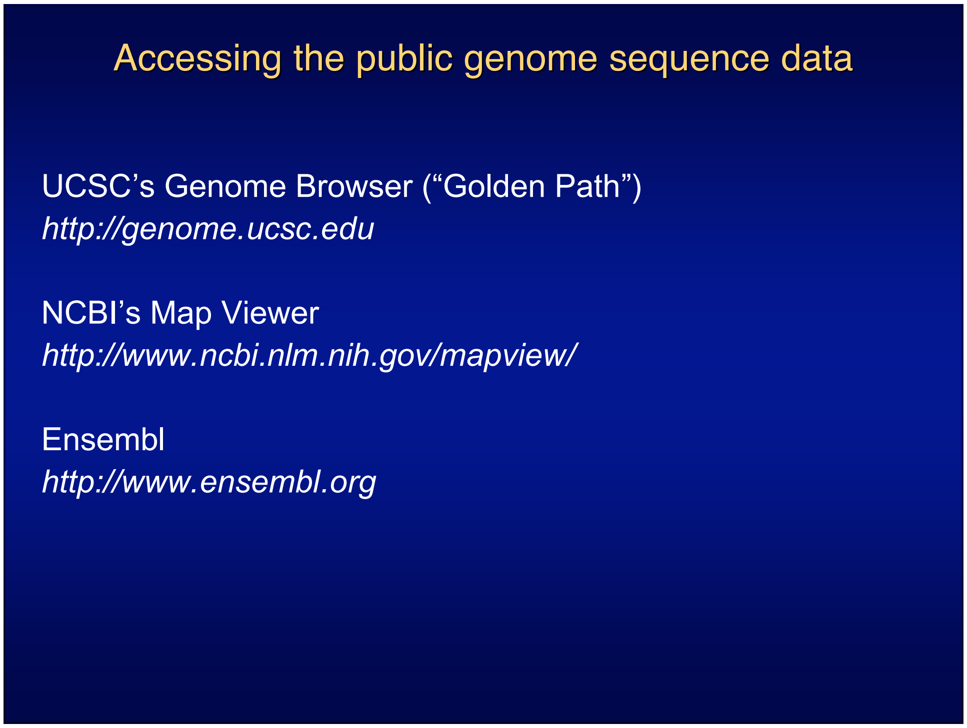


Mining Genomic Sequence Data

Tyra G. Wolfsberg, Ph.D.
NHGRI

Current Topics in Genome Analysis
January 25, 2005



Accessing the public genome sequence data

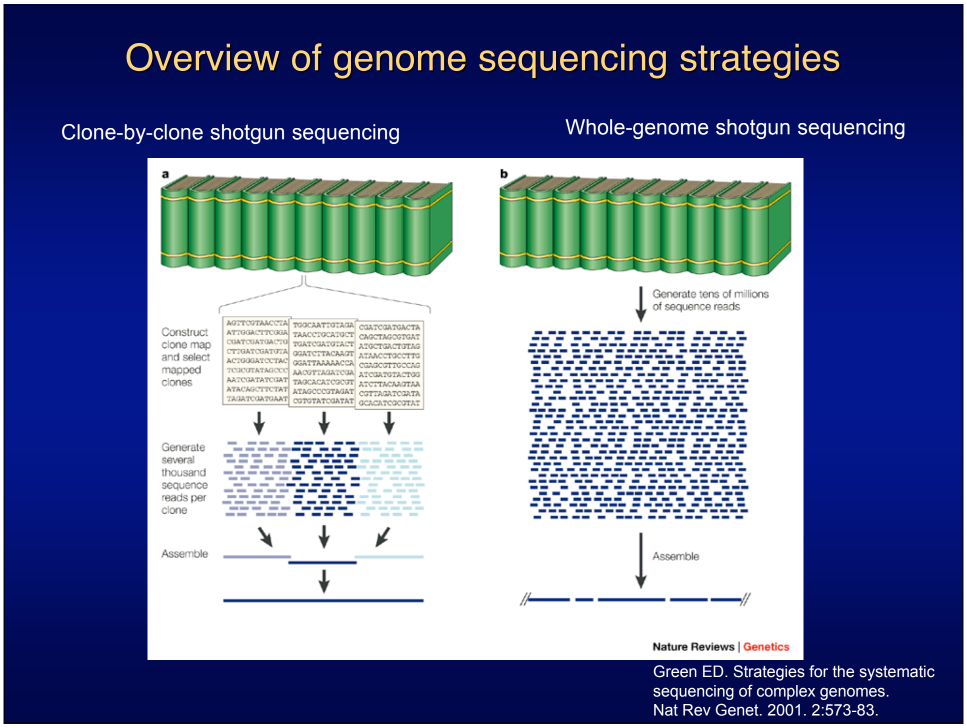
UCSC’s Genome Browser (“Golden Path”)
http://genome.ucsc.edu

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

Ensembl
http://www.ensembl.org

Types of data integrated in genome browsers

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

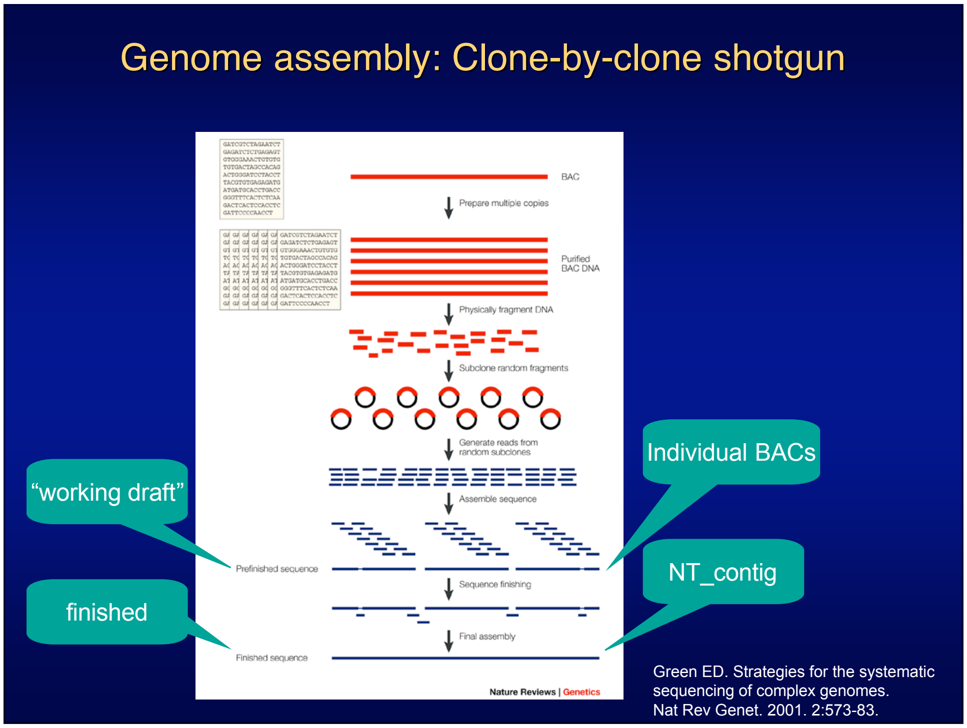


Overview of genome sequencing strategies

Clone-by-clone shotgun sequencing

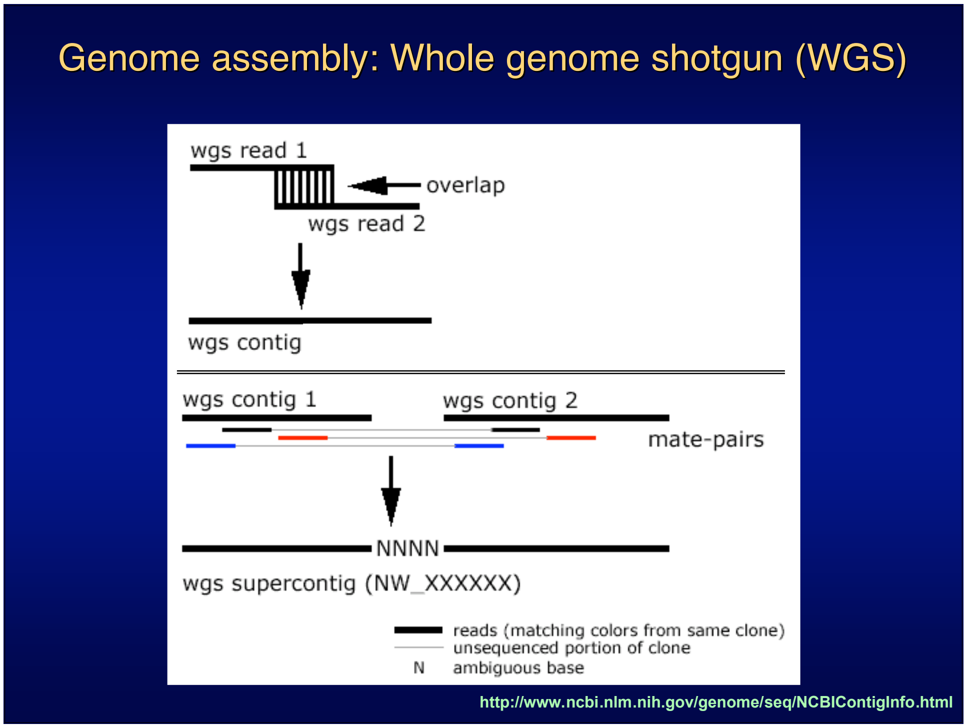
Whole-genome shotgun sequencing

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



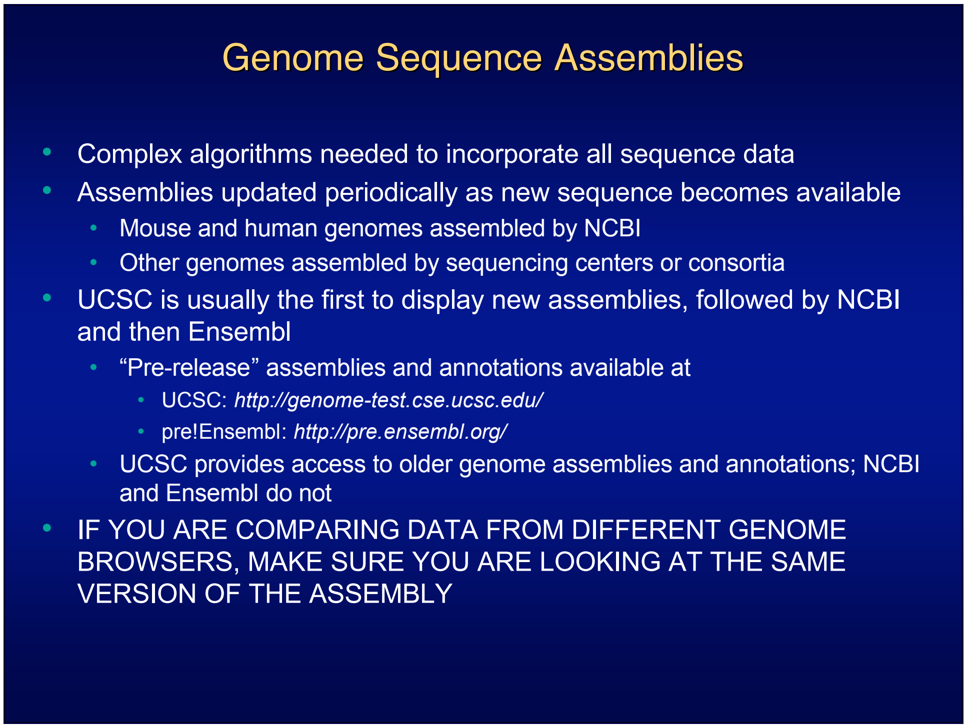
Genome assembly: Clone-by-clone shotgun

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



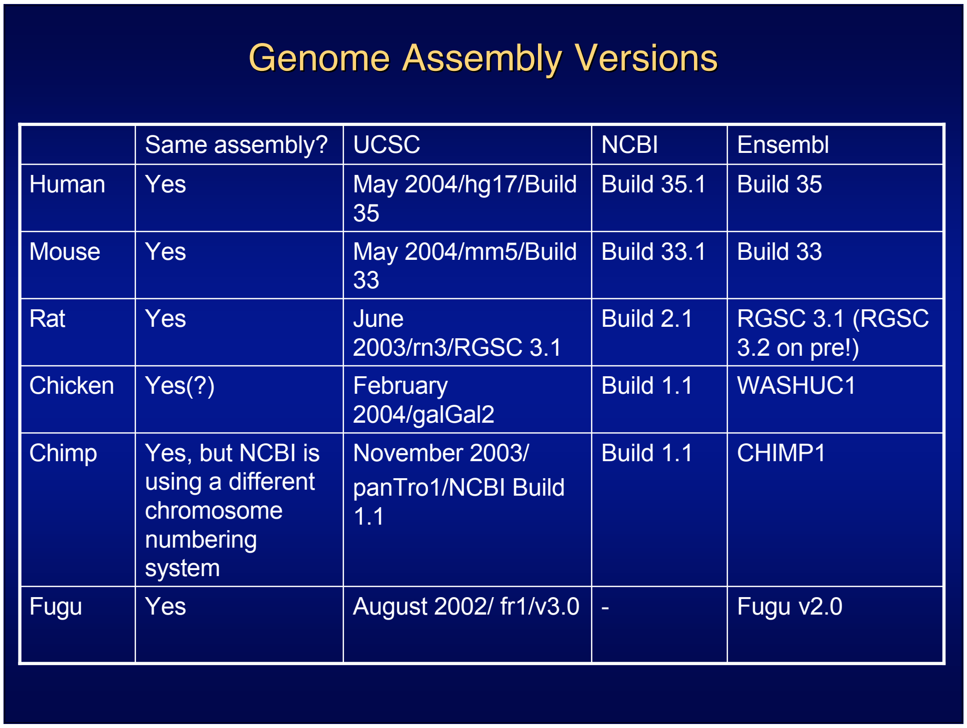
Genome assembly: Whole genome shotgun (WGS)

http://www.ncbi.nlm.nih.gov/genome/seq/NCBIContigInfo.html



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
- UCSC is usually the first to display new assemblies, followed by NCBI and then Ensembl
  - “Pre-release” assemblies and annotations available at
    - UCSC: *http://genome-test.cse.ucsc.edu/*
    - pre!Ensembl: *http://pre.ensembl.org/*
  - UCSC provides access to older genome assemblies and annotations; NCBI and Ensembl do not
- IF YOU ARE COMPARING DATA FROM DIFFERENT GENOME BROWSERS, MAKE SURE YOU ARE LOOKING AT THE SAME VERSION OF THE ASSEMBLY

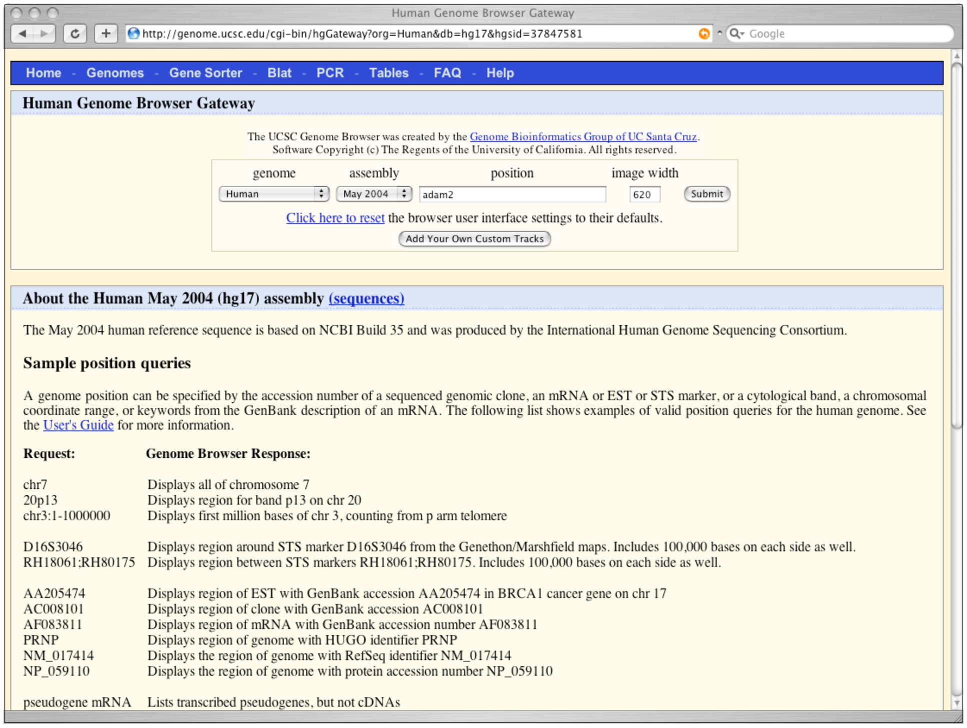


Genome Assembly Versions

	Same assembly?	UCSC	NCBI	Ensembl
Human	Yes	May 2004/hg17/Build 35	Build 35.1	Build 35
Mouse	Yes	May 2004/mm5/Build 33	Build 33.1	Build 33
Rat	Yes	June 2003/rn3/RGSC 3.1	Build 2.1	RGSC 3.1 (RGSC 3.2 on pre!)
Chicken	Yes(?)	February 2004/galGal2	Build 1.1	WASHUC1
Chimp	Yes, but NCBI is using a different chromosome numbering system	November 2003/ panTro1/NCBI Build 1.1	Build 1.1	CHIMP1
Fugu	Yes	August 2002/ fr1/v3.0	-	Fugu v2.0

UCSC

View a region in the genome by querying with a gene symbol



Human 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.

genome	assembly	position	image width
Human	May 2004	adam2	620

Click here to reset the browser user interface settings to their defaults.

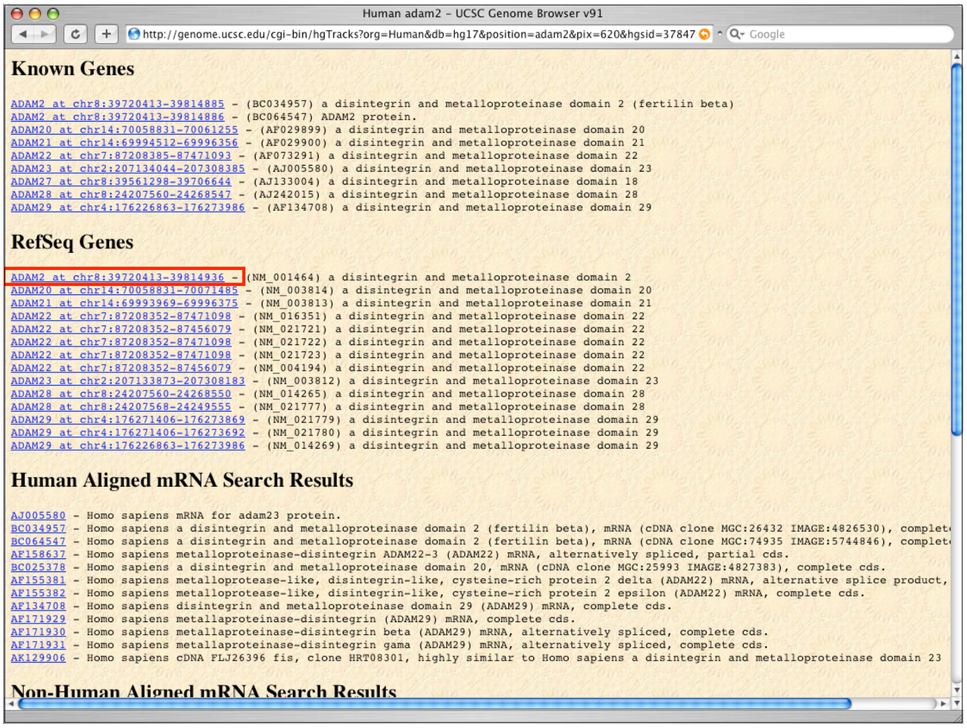
About the Human May 2004 (hg17) assembly (sequences)

The May 2004 human reference sequence is based on NCBI Build 35 and was produced by the International Human Genome Sequencing Consortium.

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, or a cytological band, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the User's Guide for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p arm telomere
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. Includes 100,000 bases on each side as well.
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 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



Known Genes

ADAM2 at chr8:39720413-39814885 - (BC034957) a disintegrin and metalloproteinase domain 2 (fertilin beta)
ADAM2 at chr8:39720413-39814886 - (BC064547) ADAM2 protein.
ADAM20 at chr14:70058831-70061255 - (AF029899) a disintegrin and metalloproteinase domain 20
ADAM21 at chr14:69994512-69996356 - (AF029900) a disintegrin and metalloproteinase domain 21
ADAM22 at chr7:87208385-87471093 - (AF073291) a disintegrin and metalloproteinase domain 22
ADAM23 at chr2:207134044-207308385 - (AJ005580) a disintegrin and metalloproteinase domain 23
ADAM27 at chr8:39561298-39706644 - (AJ133004) a disintegrin and metalloproteinase domain 18
ADAM28 at chr8:24207560-24268547 - (AJ242015) a disintegrin and metalloproteinase domain 28
ADAM29 at chr4:176226863-176273986 - (AF134708) a disintegrin and metalloproteinase domain 29

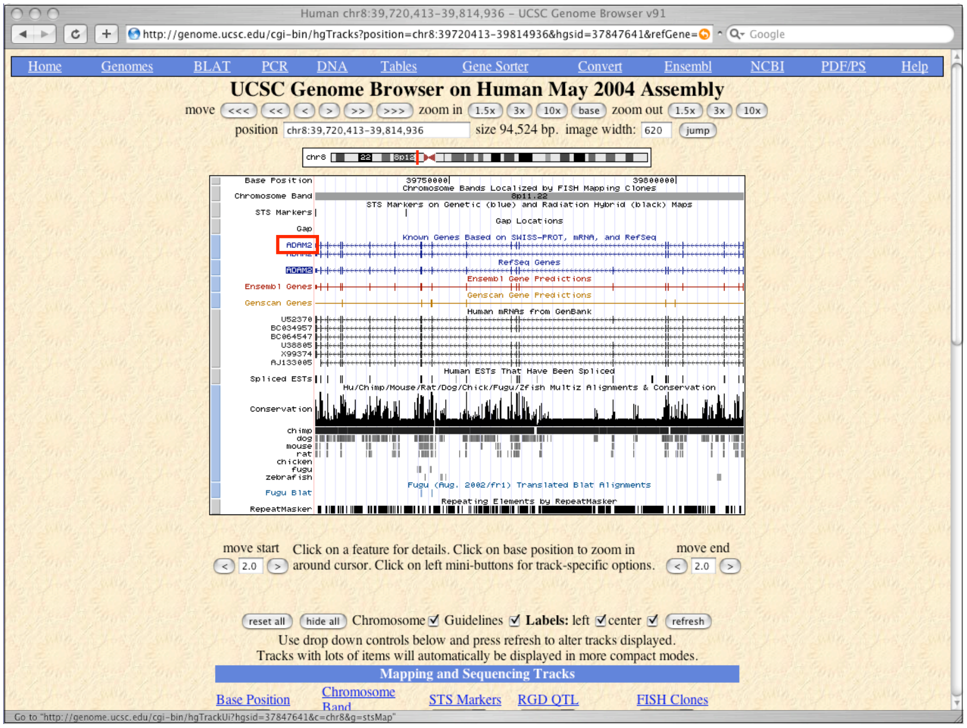
RefSeq Genes

ADAM2 at chr8:39720413-39814936 - (NM_001464) a disintegrin and metalloproteinase domain 2
ADAM20 at chr14:70058831-70071485 - (NM_003814) a disintegrin and metalloproteinase domain 20
ADAM21 at chr14:69993969-69996375 - (NM_003813) a disintegrin and metalloproteinase domain 21
ADAM22 at chr7:87208352-87471098 - (NM_016351) a disintegrin and metalloproteinase domain 22
ADAM22 at chr7:87208352-87456079 - (NM_021721) a disintegrin and metalloproteinase domain 22
ADAM22 at chr7:87208352-87471098 - (NM_021722) a disintegrin and metalloproteinase domain 22
ADAM22 at chr7:87208352-87471098 - (NM_021723) a disintegrin and metalloproteinase domain 22
ADAM22 at chr7:87208352-87456079 - (NM_004194) a disintegrin and metalloproteinase domain 22
ADAM23 at chr2:207133873-207308183 - (NM_003812) a disintegrin and metalloproteinase domain 23
ADAM28 at chr8:24207560-24268550 - (NM_014265) a disintegrin and metalloproteinase domain 28
ADAM28 at chr8:24207568-24249555 - (NM_021777) a disintegrin and metalloproteinase domain 28
ADAM29 at chr4:176271406-176273869 - (NM_021779) a disintegrin and metalloproteinase domain 29
ADAM29 at chr4:176271406-176273692 - (NM_021780) a disintegrin and metalloproteinase domain 29
ADAM29 at chr4:176226863-176273986 - (NM_014269) a disintegrin and metalloproteinase domain 29

Human Aligned mRNA Search Results

AJ005580 - Homo sapiens mRNA for adam23 protein.
BC034957 - Homo sapiens a disintegrin and metalloproteinase domain 2 (fertilin beta), mRNA (cDNA clone MGC:26432 IMAGE:4826530), complet
BC064547 - Homo sapiens a disintegrin and metalloproteinase domain 2 (fertilin beta), mRNA (cDNA clone MGC:74935 IMAGE:5744846), complet
AF158637 - Homo sapiens metalloproteinase-disintegrin ADAM22-3 (ADAM22) mRNA, alternatively spliced, partial cds.
BC025378 - Homo sapiens a disintegrin and metalloproteinase domain 20, mRNA (cDNA clone MGC:25993 IMAGE:4827383), complete cds.
AF155381 - Homo sapiens metalloprotease-like, disintegrin-like, cysteine-rich protein 2 delta (ADAM22) mRNA, alternative splice product,
AF155382 - Homo sapiens metalloprotease-like, disintegrin-like, cysteine-rich protein 2 epsilon (ADAM22) mRNA, complete cds.
AF134708 - Homo sapiens disintegrin and metalloproteinase domain 29 (ADAM29) mRNA, complete cds.
AF171929 - Homo sapiens metallaproteinase-disintegrin (ADAM29) mRNA, complete cds.
AF171930 - Homo sapiens metallaproteinase-disintegrin beta (ADAM29) mRNA, alternatively spliced, complete cds.
AF171931 - Homo sapiens metallaproteinase-disintegrin gama (ADAM29) mRNA, alternatively spliced, complete cds.
AK129906 - Homo sapiens cDNA FLJ26396 fis, clone HRT08301, highly similar to Homo sapiens a disintegrin and metalloproteinase domain 23

Non-Human Aligned mRNA Search Results



UCSC Genome Browser on Human May 2004 Assembly

position chr8:39,720,413-39,814,936 size 94,524 bp. image width: 620

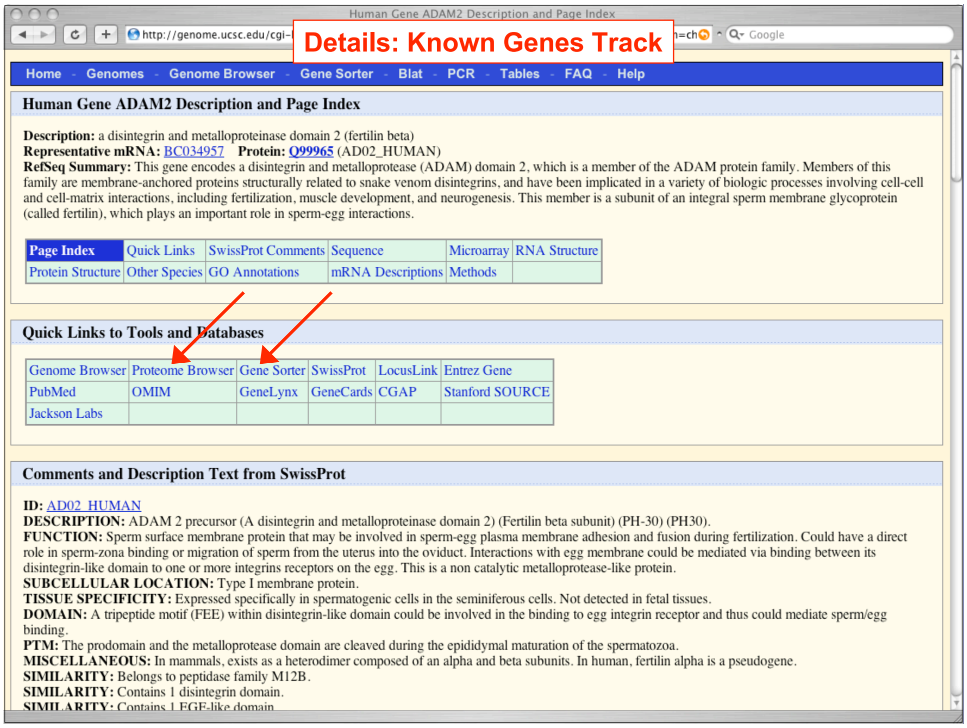
Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.

move start 2.0 move end 2.0

Chromosome ☑ Guidelines ☑ **Labels:** left ☑ center ☑

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.

Go to "http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=37847641&c=chr8&g=stsMap"



Details: Known Genes Track

Human Gene ADAM2 Description and Page Index

Description: a disintegrin and metalloproteinase domain 2 (fertilin beta)
Representative mRNA: BC034957 **Protein:** **Q99965** (AD02_HUMAN)
RefSeq Summary: This gene encodes a disintegrin and metalloprotease (ADAM) domain 2, which is a member of the ADAM protein family. Members of this family are membrane-anchored proteins structurally related to snake venom disintegrins, and have been implicated in a variety of biologic 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.

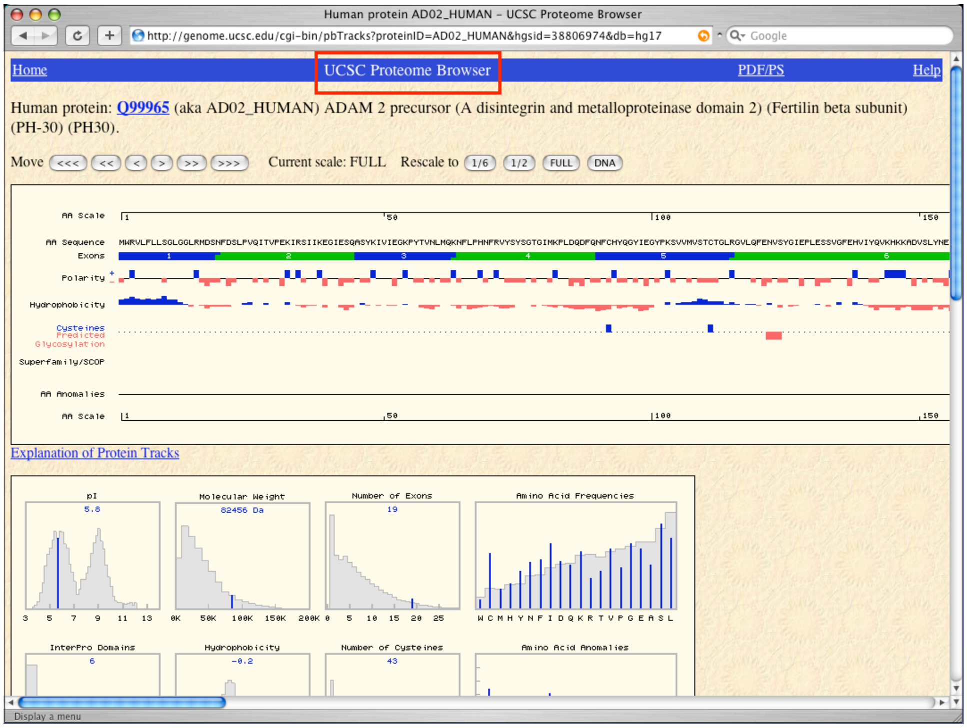
Page Index	Quick Links	SwissProt Comments	Sequence	Microarray	RNA Structure
Protein Structure	Other Species	GO Annotations	mRNA Descriptions	Methods	

Quick Links to Tools and Databases

Genome Browser	Proteome Browser	Gene Sorter	SwissProt	LocusLink	Entrez Gene
PubMed	OMIM	GeneLynx	GeneCards	CGAP	Stanford SOURCE
Jackson Labs					

Comments and Description Text from SwissProt

ID: AD02_HUMAN
DESCRIPTION: ADAM 2 precursor (A disintegrin and metalloproteinase domain 2) (Fertilin beta subunit) (PH-30) (PH30).
FUNCTION: Sperm surface membrane protein that may be involved in sperm-egg plasma membrane adhesion and fusion during fertilization. Could have a direct role in sperm-zona binding or migration of sperm from the uterus into the oviduct. Interactions with egg membrane could be mediated via binding between its disintegrin-like domain to one or more integrins receptors on the egg. This is a non catalytic metalloprotease-like protein.
SUBCELLULAR LOCATION: Type I membrane protein.
TISSUE SPECIFICITY: Expressed specifically in spermatogenic cells in the seminiferous cells. Not detected in fetal tissues.
DOMAIN: A tripeptide motif (FEE) within disintegrin-like domain could be involved in the binding to egg integrin receptor and thus could mediate sperm/egg binding.
PTM: The prodomain and the metalloprotease domain are cleaved during the epididymal maturation of the spermatozoa.
MISCELLANEOUS: In mammals, exists as a heterodimer composed of an alpha and beta subunits. In human, fertilin alpha is a pseudogene.
SIMILARITY: Belongs to peptidase family M12B.
SIMILARITY: Contains 1 disintegrin domain.
SIMILARITY: Contains 1 EGF-like domain.

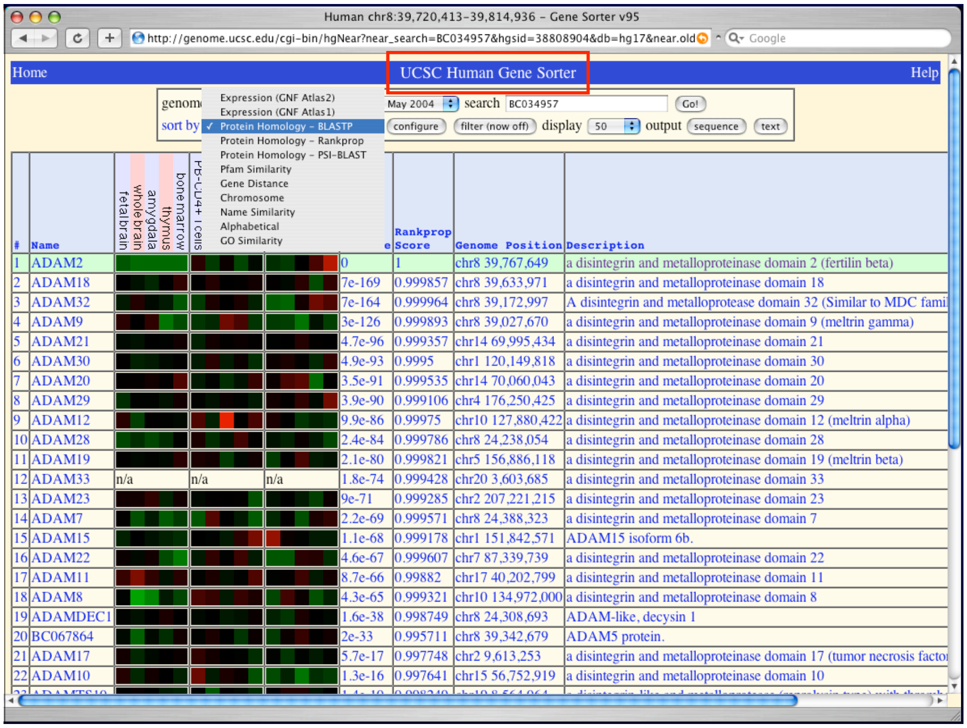


UCSC Proteome Browser

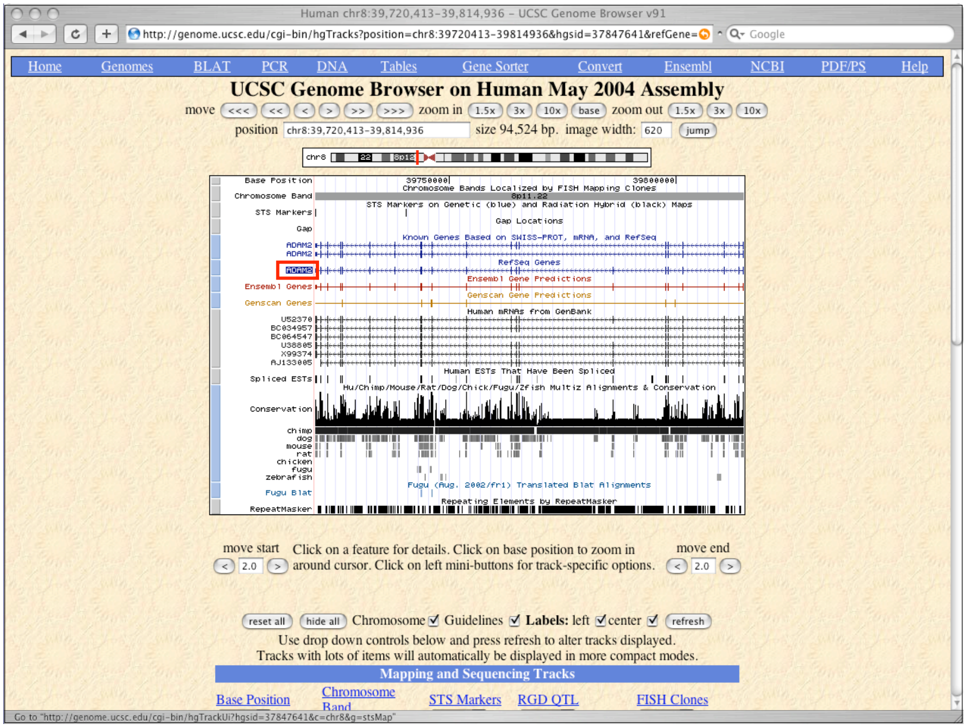
Human protein: **Q99965** (aka AD02_HUMAN) ADAM 2 precursor (A disintegrin and metalloproteinase domain 2) (Fertilin beta subunit) (PH-30) (PH30).

Move <<< << < > >> >>> Current scale: FULL Rescale to 1/6 1/2 FULL DNA

Explanation of Protein Tracks



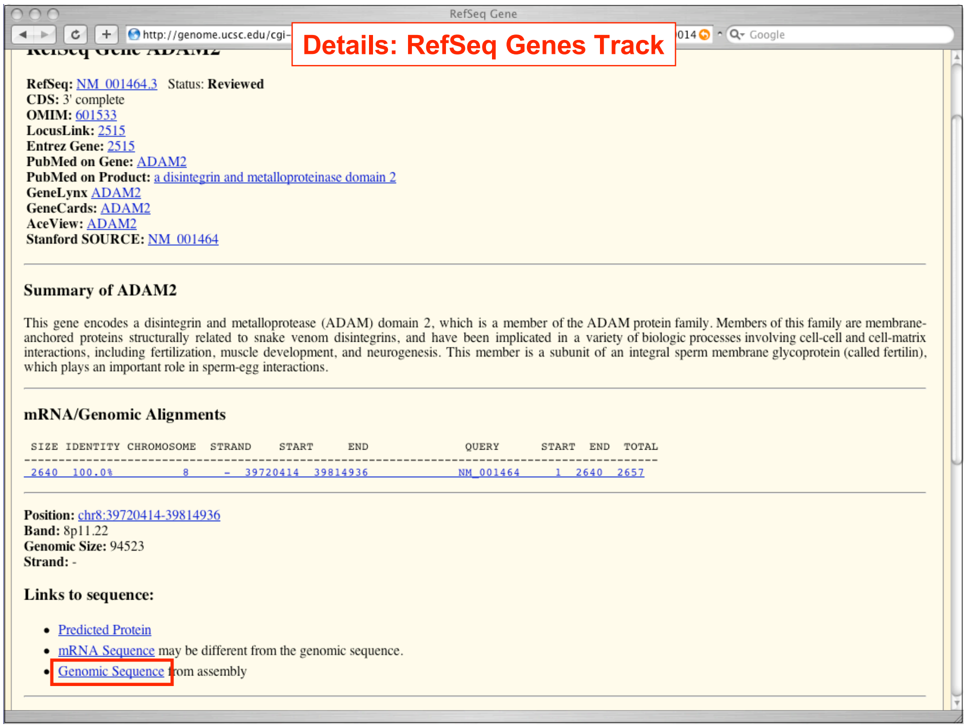
#	Name	fetalbrain wholebrain amygdala thymus bonemarrow	PB-CD4+ T cells		e	Rankprop Score	Genome Position	Description
1	ADAM2				0	1	chr8 39,767,649	a disintegrin and metalloproteinase domain 2 (fertilin beta)
2	ADAM18				7e-169	0.999857	chr8 39,633,971	a disintegrin and metalloproteinase domain 18
3	ADAM32				7e-164	0.999964	chr8 39,172,997	A disintegrin and metalloprotease domain 32 (Similar to MDC fami
4	ADAM9				3e-126	0.999893	chr8 39,027,670	a disintegrin and metalloproteinase domain 9 (meltrin gamma)
5	ADAM21				4.7e-96	0.999357	chr14 69,995,434	a disintegrin and metalloproteinase domain 21
6	ADAM30				4.9e-93	0.9995	chr1 120,149,818	a disintegrin and metalloproteinase domain 30
7	ADAM20				3.5e-91	0.999535	chr14 70,060,043	a disintegrin and metalloproteinase domain 20
8	ADAM29				3.9e-90	0.999106	chr4 176,250,425	a disintegrin and metalloproteinase domain 29
9	ADAM12				9.9e-86	0.99975	chr10 127,880,422	a disintegrin and metalloproteinase domain 12 (meltrin alpha)
10	ADAM28				2.4e-84	0.999786	chr8 24,238,054	a disintegrin and metalloproteinase domain 28
11	ADAM19				2.1e-80	0.999821	chr5 156,886,118	a disintegrin and metalloproteinase domain 19 (meltrin beta)
12	ADAM33	n/a	n/a	n/a	1.8e-74	0.999428	chr20 3,603,685	a disintegrin and metalloproteinase domain 33
13	ADAM23				9e-71	0.999285	chr2 207,221,215	a disintegrin and metalloproteinase domain 23
14	ADAM7				2.2e-69	0.999571	chr8 24,388,323	a disintegrin and metalloproteinase domain 7
15	ADAM15				1.1e-68	0.999178	chr1 151,842,571	ADAM15 isoform 6b.
16	ADAM22				4.6e-67	0.999607	chr7 87,339,739	a disintegrin and metalloproteinase domain 22
17	ADAM11				8.7e-66	0.99882	chr17 40,202,799	a disintegrin and metalloproteinase domain 11
18	ADAM8				4.3e-65	0.999321	chr10 134,972,000	a disintegrin and metalloproteinase domain 8
19	ADAMDEC1				1.6e-38	0.998749	chr8 24,308,693	ADAM-like, decysin 1
20	BC067864				2e-33	0.995711	chr8 39,342,679	ADAM5 protein.
21	ADAM17				5.7e-17	0.997748	chr2 9,613,253	a disintegrin and metalloproteinase domain 17 (tumor necrosis facto
22	ADAM10				1.3e-16	0.997641	chr15 56,752,919	a disintegrin and metalloproteinase domain 10



UCSC Genome Browser on Human May 2004 Assembly

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

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.



Details: RefSeq Genes Track

RefSeq Gene ADAM2

RefSeq: NM_001464.3 Status: **Reviewed**
CDS: 3' complete
OMIM: 601533
LocusLink: 2515
Entrez Gene: 2515
PubMed on Gene: ADAM2
PubMed on Product: a disintegrin and metalloproteinase domain 2
GeneLynx ADAM2
GeneCards: ADAM2
AceView: ADAM2
Stanford SOURCE: NM_001464

Summary of ADAM2

This gene encodes a disintegrin and metalloprotease (ADAM) domain 2, which is a member of the ADAM protein family. Members of this family are membrane-anchored proteins structurally related to snake venom disintegrins, and have been implicated in a variety of biologic processes involving cell-cell and cell-matrix interactions, including fertilization, muscle development, and neurogenesis. This member is a subunit of an integral sperm membrane glycoprotein (called fertilin), which plays an important role in sperm-egg interactions.

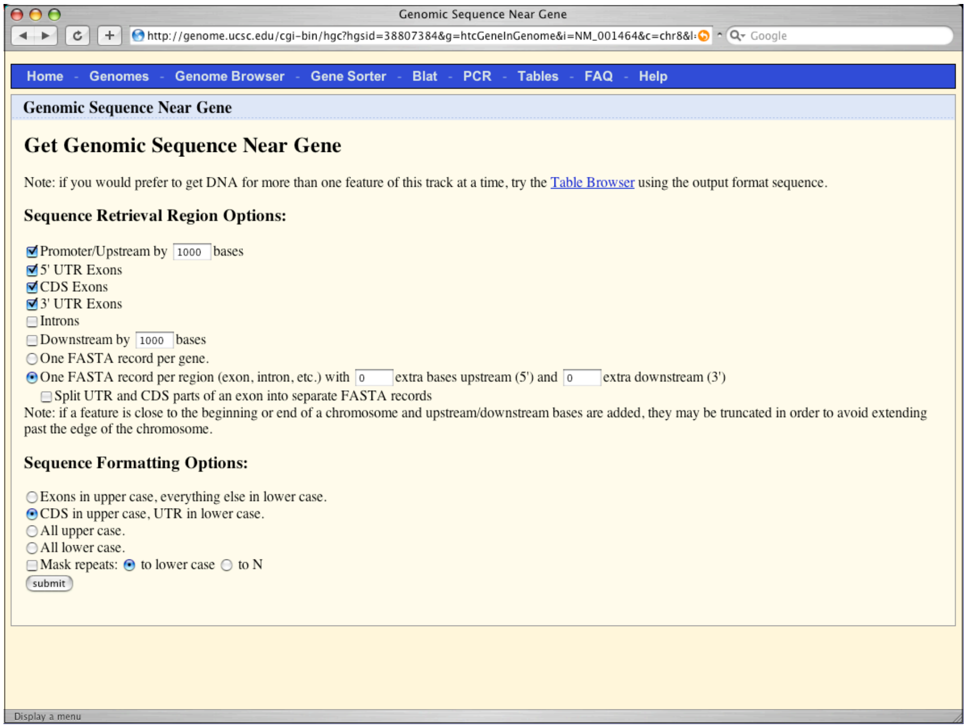
mRNA/Genomic Alignments

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

Position: chr8:39720414-39814936
Band: 8p11.22
Genomic Size: 94523
Strand: -

Links to sequence:

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



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.

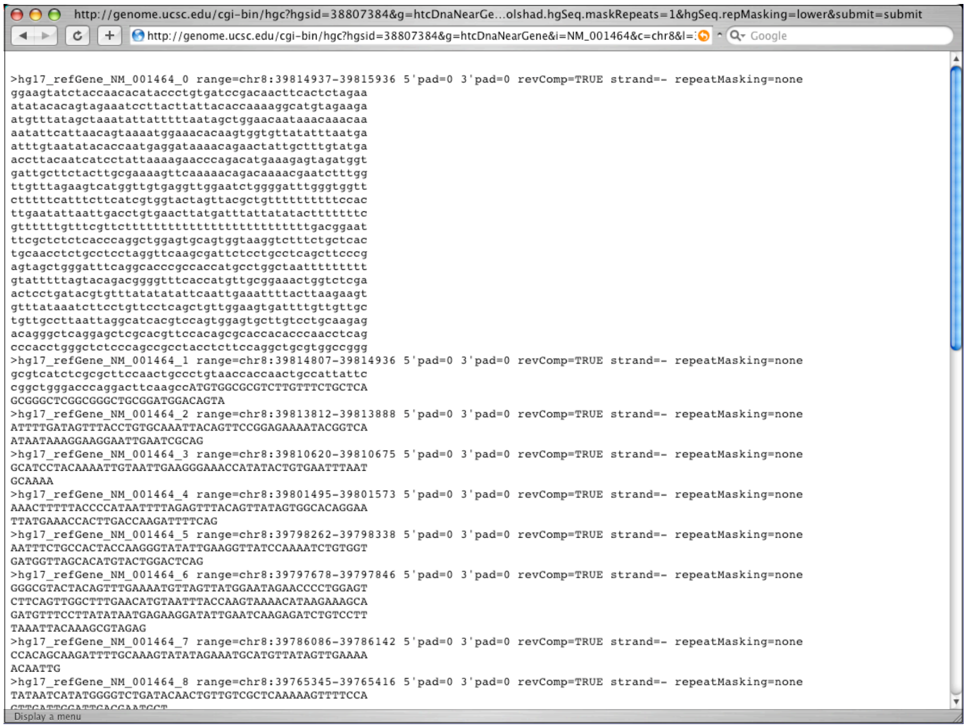
Sequence Retrieval Region Options:

☑ Promoter/Upstream by 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:

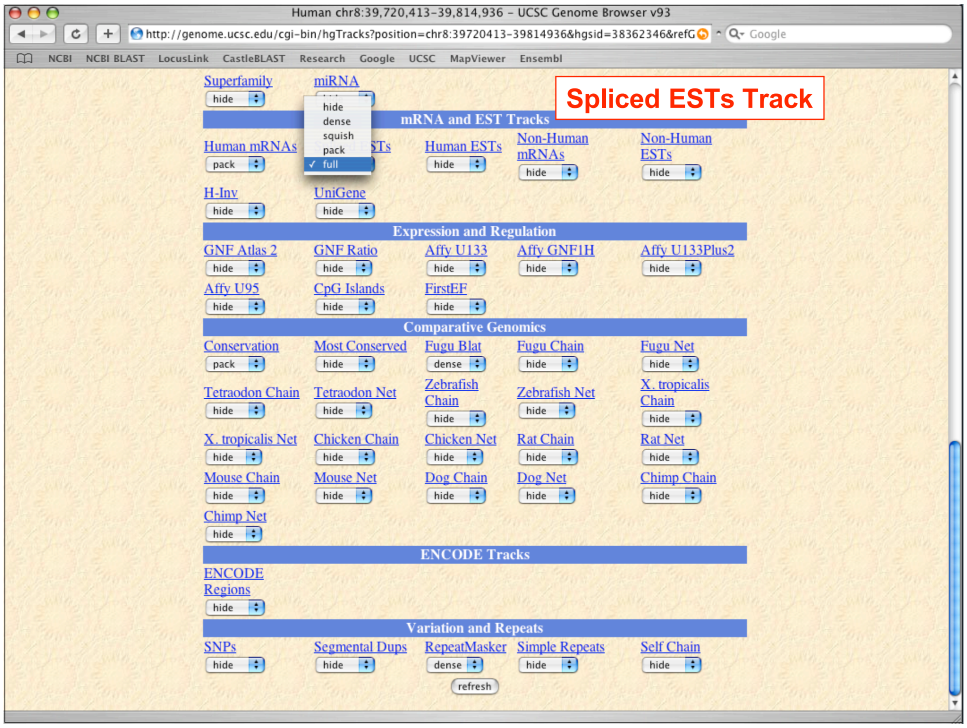
○ Exons in upper case, everything else in lower case.
● CDS in upper case, UTR in lower case.
○ All upper case.
○ All lower case.
☐ Mask repeats: ● to lower case ○ to N
submit



```
>hg17_refGene_NM_001464_0 range=chr8:39814937-39815936 5'pad=0 3'pad=0 revComp=TRUE strand=- repeatMasking=none
ggaagtatctaccaacacataccctgtgatccgacaacttcactctagaa
atatacacagtagaaatccttacttattacaccaaaaggcatgtagaaga
atgtttatagctaaatattatttttaatagctggaacaataaacaaacaa
aatattcattaacagtaaaatggaaacacaagtggtgttatatttaatga
atttgtaatatacaccaatgaggataaaacagaactattgctttgtatga
accttacaatcatcctattaaaagaacccagacatgaaagagtagatggt
gattgcttctacttgcgaaaagttcaaaaacagacaaaacgaatctttgg
ttgtttagaagtcatggttgtgaggttggaatctggggatttgggtggtt
ctttttcatttcttcatcgtggtactagttacgctgttttttttttccac
ttgaatattaattgacctgtgaacttatgatttattatatactttttttc
gtttttgtttcgttctttttttttttttttttttttttttgacggaat
ttcgctctctcacccaggctggagtgcagtggtaaggtctttctgctcac
tgcaacctctgcctcctaggttcaagcgattctcctgcctcagcttcccg
agtagctgggatttcaggcacccgccaccatgcctggctaatttttttttt
gtatttttagtacagacggggtttcaccatgttgcggaaactggtctcga
actcctgatacgtgtttatatatattcaattgaaattttacttaagaagt
gtttataaatcttcctgttcctcagctgttggaagtgattttgttgttgc
tgttgccttaattaggcatcacgtccagtggagtgcttgtcctgcaagag
acagggctcaggagctcgcacgttccacagcgcaccacacccaacctcag
cccacctgggctctcccagccgcctacctcttccaggctgcgtggccggg
>hg17_refGene_NM_001464_1 range=chr8:39814807-39814936 5'pad=0 3'pad=0 revComp=TRUE strand=- repeatMasking=none
gcgtcatctcgcgcttccaactgccctgtaaccaccaactgccattattc
cggctgggacccaggacttcaagccATGTGGCGCGTCTTGTTTCTGCTCA
GCGGGCTCGGCGGGCTGCGGATGGACAGTA
>hg17_refGene_NM_001464_2 range=chr8:39813812-39813888 5'pad=0 3'pad=0 revComp=TRUE strand=- repeatMasking=none
ATTTTGATAGTTTACCTGTGCAAATTACAGTTCCGGAGAAAATACGGTCA
ATAATAAAGGAAGGAATTGAATCGCAG
>hg17_refGene_NM_001464_3 range=chr8:39810620-39810675 5'pad=0 3'pad=0 revComp=TRUE strand=- repeatMasking=none
GCATCCTACAAAATTGTAATTGAAGGGAAACCATATACTGTGAATTTAAT
GCAAAA
>hg17_refGene_NM_001464_4 range=chr8:39801495-39801573 5'pad=0 3'pad=0 revComp=TRUE strand=- repeatMasking=none
AAACTTTTTACCCCATAATTTTAGAGTTTACAGTTATAGTGGCACAGGAA
TTATGAAACCACTTGACCAAGATTTTCAG
>hg17_refGene_NM_001464_5 range=chr8:39798262-39798338 5'pad=0 3'pad=0 revComp=TRUE strand=- repeatMasking=none
AATTTCTGCCACTACCAAGGGTATATTGAAGGTTATCCAAAATCTGTGGT
GATGGTTAGCACATGTACTGGACTCAG
>hg17_refGene_NM_001464_6 range=chr8:39797678-39797846 5'pad=0 3'pad=0 revComp=TRUE strand=- repeatMasking=none
GGGCGTACTACAGTTTGAAAATGTTAGTTATGGAATAGAACCCCTGGAGT
CTTCAGTTGGCTTTGAACATGTAATTTACCAAGTAAAACATAAGAAAGCA
GATGTTTCCTTATATAATGAGAAGGATATTGAATCAAGAGATCTGTCCTT
TAAATTACAAAGCGTAGAG
>hg17_refGene_NM_001464_7 range=chr8:39786086-39786142 5'pad=0 3'pad=0 revComp=TRUE strand=- repeatMasking=none
CCACAGCAAGATTTTGCAAAGTATATAGAAATGCATGTTATAGTTGAAAA
ACAATTG
>hg17_refGene_NM_001464_8 range=chr8:39765345-39765416 5'pad=0 3'pad=0 revComp=TRUE strand=- repeatMasking=none
TATAATCATATGGGGTCTGATACAACTGTTGTCGCTCAAAAAGTTTTCCA
```

UCSC

Change the tracks displayed on the Genome Browser



Human chr8:39,720,413-39,814,936 - UCSC Genome Browser v93

http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr8:39720413-39814936&hgsid=38362346&refG

Superfamily: hide
miRNA: full

mRNA and EST Tracks

Human mRNAs: pack
Human ESTs: hide
Non-Human mRNAs: hide
Non-Human ESTs: hide
H-Inv: hide
UniGene: hide

Expression and Regulation

GNF Atlas 2: hide
GNF Ratio: hide
Affy U133: hide
Affy GNF1H: hide
Affy U133Plus2: hide
Affy U95: hide
CpG Islands: hide
FirstEF: hide

Comparative Genomics

Conservation: pack
Most Conserved: hide
Fugu Blat: dense
Fugu Chain: hide
Fugu Net: hide
Tetraodon Chain: hide
Tetraodon Net: hide
Zebrafish Chain: hide
Zebrafish Net: hide
X. tropicalis Chain: hide
X. tropicalis Net: hide
Chicken Chain: hide
Chicken Net: hide
Rat Chain: hide
Rat Net: hide
Mouse Chain: hide
Mouse Net: hide
Dog Chain: hide
Dog Net: hide
Chimp Chain: hide
Chimp Net: hide

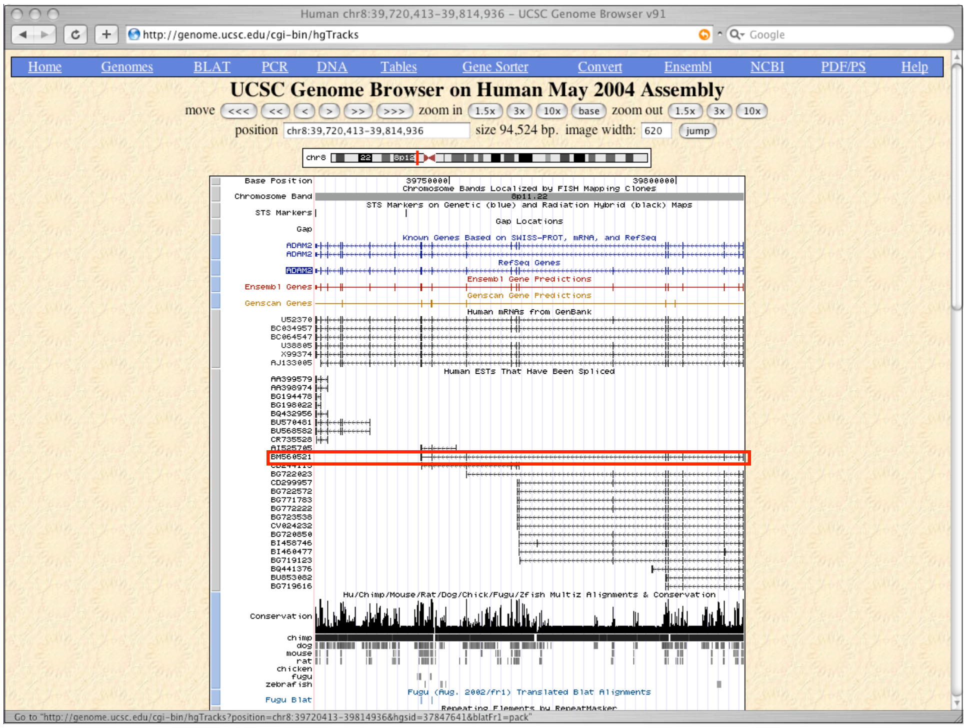
ENCODE Tracks

ENCODE Regions: hide

Variation and Repeats

SNPs: hide
Segmental Dups: hide
RepeatMasker: dense
Simple Repeats: hide
Self Chain: hide

Spliced ESTs Track



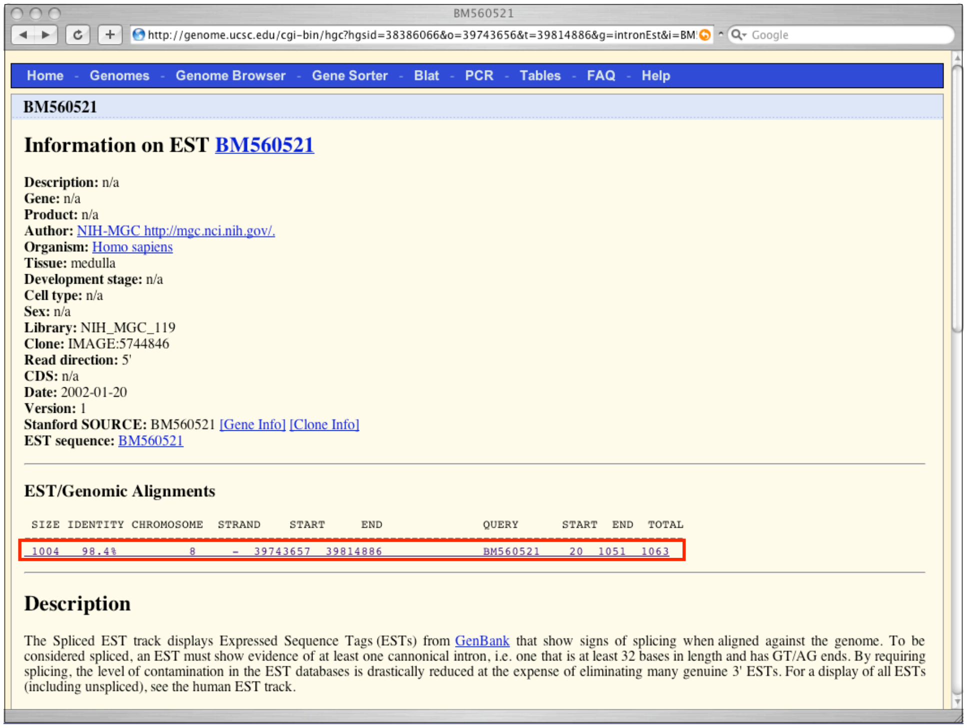
Human chr8:39,720,413-39,814,936 - UCSC Genome Browser v91

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

UCSC Genome Browser on Human May 2004 Assembly

position chr8:39,720,413-39,814,936 size 94,524 bp. image width: 620

Go to "http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr8:39720413-39814936&hgsid=37847641&blatFr1=pack"



BM560521

Information on EST BM560521

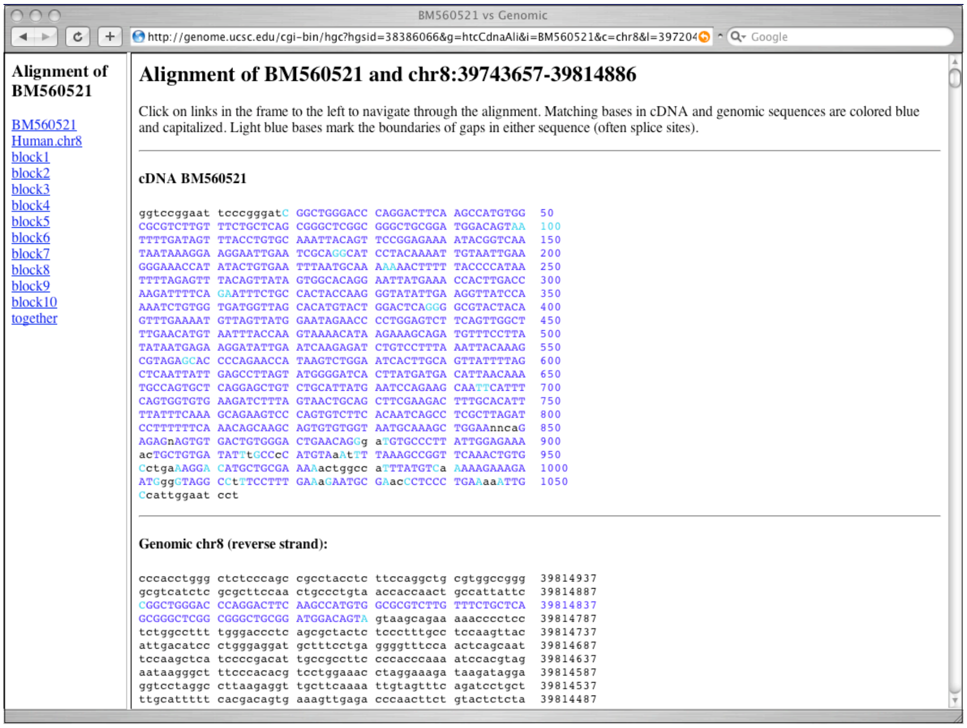
Description: n/a
Gene: n/a
Product: n/a
Author: NIH-MGC http://mgc.nci.nih.gov/.
Organism: Homo sapiens
Tissue: medulla
Development stage: n/a
Cell type: n/a
Sex: n/a
Library: NIH_MGC_119
Clone: IMAGE:5744846
Read direction: 5'
CDS: n/a
Date: 2002-01-20
Version: 1
Stanford SOURCE: BM560521 [Gene Info] [Clone Info]
EST sequence: BM560521

EST/Genomic Alignments

```
 SIZE IDENTITY CHROMOSOME  STRAND    START     END              QUERY      START  END  TOTAL
 -----------------------------------------------------------------------------------------
 1004   98.4%           8      -   39743657  39814886           BM560521     20  1051  1063
```

Description

The Spliced EST track displays Expressed Sequence Tags (ESTs) from GenBank that show signs of splicing when aligned against the genome. To be considered spliced, an EST must show evidence of at least one cannonical intron, i.e. one that is at least 32 bases in length and has GT/AG ends. By requiring splicing, the level of contamination in the EST databases is drastically reduced at the expense of eliminating many genuine 3' ESTs. For a display of all ESTs (including unspliced), see the human EST track.



Alignment of BM560521

BM560521
Human.chr8
block1
block2
block3
block4
block5
block6
block7
block8
block9
block10
together

Alignment of BM560521 and chr8:39743657-39814886

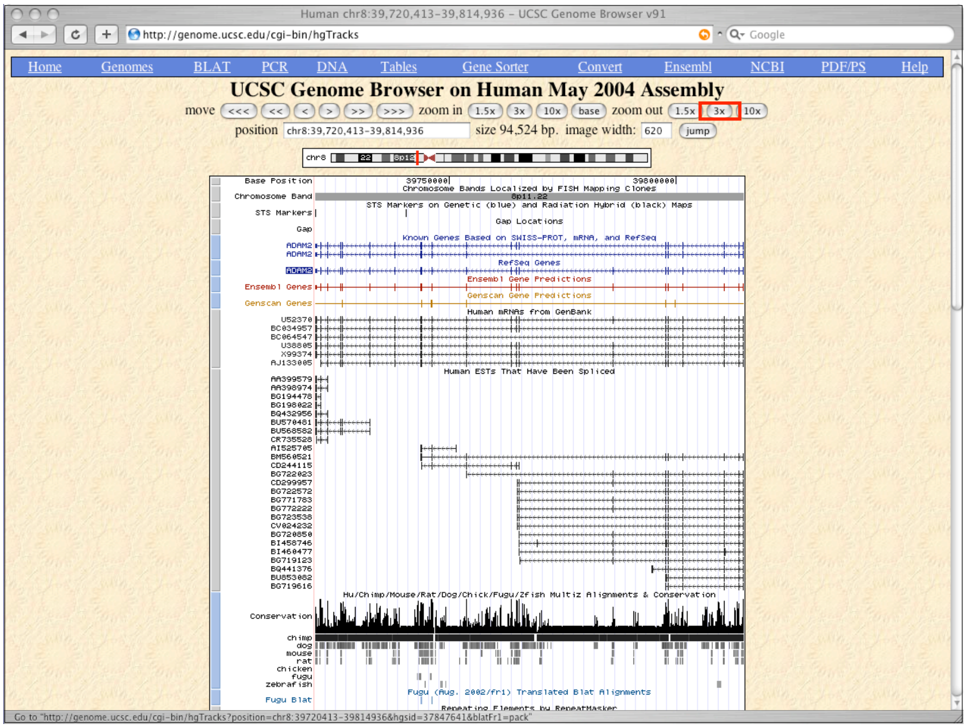
Click on links in the frame to the left to navigate through the alignment. Matching bases in cDNA and genomic sequences are colored blue and capitalized. Light blue bases mark the boundaries of gaps in either sequence (often splice sites).

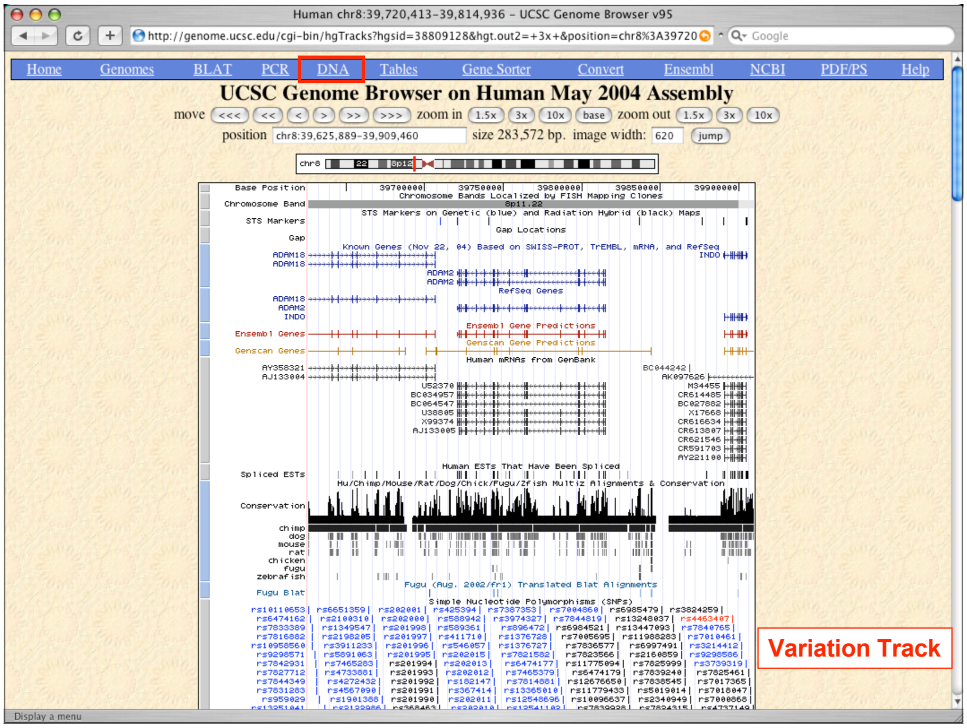
cDNA BM560521

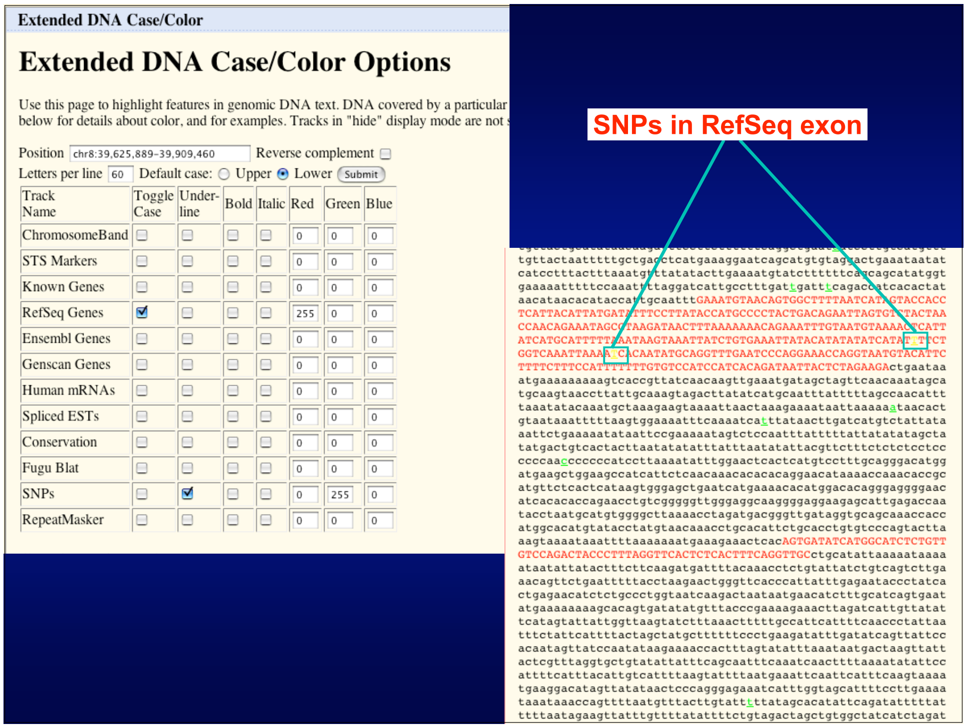
```
ggtccggaat tcccgggatC GGCTGGGACC CAGGACTTCA AGCCATGTGG  50
CGCGTCTTGT TTCTGCTCAG CGGGCTCGGC GGGCTGCGGA TGGACAGTAA  100
TTTTGATAGT TTACCTGTGC AAATTACAGT TCCGGAGAAA ATACGGTCAA  150
TAATAAAGGA AGGAATTGAA TCGCAGGCAT CCTACAAAAT TGTAATTGAA  200
GGGAAACCAT ATACTGTGAA TTTAATGCAA AAAAACTTTT TACCCCATAA  250
TTTTAGAGTT TACAGTTATA GTGGCACAGG AATTATGAAA CCACTTGACC  300
AAGATTTTCA GAATTTCTGC CACTACCAAG GGTATATTGA AGGTTATCCA  350
AAATCTGTGG TGATGGTTAG CACATGTACT GGACTCAGGG GCGTACTACA  400
GTTTGAAAAT GTTAGTTATG GAATAGAACC CCTGGAGTCT TCAGTTGGCT  450
TTGAACATGT AATTTACCAA GTAAAACATA AGAAAGCAGA TGTTTCCTTA  500
TATAATGAGA AGGATATTGA ATCAAGAGAT CTGTCCTTTA AATTACAAAG  550
CGTAGAGCAC CCCAGAACCA TAAGTCTGGA ATCACTTGCA GTTATTTTAG  600
CTCAATTATT GAGCCTTAGT ATGGGGATCA CTTATGATGA CATTAACAAA  650
TGCCAGTGCT CAGGAGCTGT CTGCATTATG AATCCAGAAG CAATTCATTT  700
CAGTGGTGTG AAGATCTTTA GTAACTGCAG CTTCGAAGAC TTTGCACATT  750
TTATTTCAAA GCAGAAGTCC CAGTGTCTTC ACAATCAGCC TCGCTTAGAT  800
CCTTTTTTCA AACAGCAAGC AGTGTGTGGT AATGCAAAGC TGGAAnncaG  850
AGAGnAGTGT GACTGTGGGA CTGAACAGGg aTGTGCCCTT ATTGGAGAAA  900
acTGCTGTGA TATTtGCCcC ATGTAaAtTT TAAAGCCGGT TCAAACTGTG  950
CctgaAAGGA CATGCTGCGA AAAactggcc aTTTATGTCa AAAAGAAAGA  1000
ATGggGTAGG CCtTTCCTTT GAAaGAATGC GAacCCTCCC TGAAaaATTG  1050
Ccattggaat cct
```

Genomic chr8 (reverse strand):

```
cccacctggg ctctcccagc cgcctacctc ttccaggctg cgtggccggg  39814937
gcgtcatctc gcgcttccaa ctgccctgta accaccaact gccattattc  39814887
CGGCTGGGAC CCAGGACTTC AAGCCATGTG GCGCGTCTTG TTTCTGCTCA  39814837
GCGGGCTCGG CGGGCTGCGG ATGGACAGTA gtaagcagaa aaacccctcc  39814787
tctggccttt tgggaccctc agcgctactc tccctttgcc tccaagttac  39814737
attgacatcc ctgggaggat gctttcctga ggggtttcca actcagcaat  39814687
tccaagctca tccccgacat tgccgccttc cccacccaaa atccacgtag  39814637
aataagggct ttcccacacg tcctggaaac ctaggaaaga taagatagga  39814587
ggtcctaggc cttaagaggt tgcttcaaaa ttgtagtttc agatcctgct  39814537
ttgcattttt cacgacagtg aaagttgaga cccaacttct gtactctcta  39814487
```


Human chr8:39,720,413-39,814,936 - UCSC Genome Browser v91
http://genome.ucsc.edu/cgi-bin/hgTracks
Home Genomes BLAT PCR DNA Tables Gene Sorter Convert Ensembl NCBI PDF/PS Help
UCSC Genome Browser on Human May 2004 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position chr8:39,720,413-39,814,936 size 94,524 bp. image width: 620 jump
Base Position
Chromosome Band
STS Markers
Gap
Known Genes Based on SWISS-PROT, TrEMBL, mRNA, and RefSeq
RefSeq Genes
Ensembl Genes
Genscan Genes
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
Hu/Chimp/Mouse/Rat/Dog/Chick/Fugu/Zfish Multiz Alignments & Conservation
Conservation
chimp
dog
mouse
rat
chicken
fugu
zebrafish
Fugu Blat
Fugu (Aug. 2002/fr1) Translated Blat Alignments
Repeating Elements by RepeatMasker
Go to "http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr8:39720413-39814936&hgsid=37847641&blatFr1=pack"


Human chr8:39,720,413-39,814,936 - UCSC Genome Browser v95
http://genome.ucsc.edu/cgi-bin/hgTracks?hgsid=38809128&hgt.out2=+3x+&position=chr8%3A39720
Home Genomes BLAT PCR DNA Tables Gene Sorter Convert Ensembl NCBI PDF/PS Help
UCSC Genome Browser on Human May 2004 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position chr8:39,625,889-39,909,460 size 283,572 bp. image width: 620 jump
Base Position
Chromosome Band
STS Markers
Gap
Known Genes (Nov 22, 04) Based on SWISS-PROT, TrEMBL, mRNA, and RefSeq
RefSeq Genes
Ensembl Genes
Genscan Genes
Human mRNAs from GenBank
Spliced ESTs
Human ESTs That Have Been Spliced
Hu/Chimp/Mouse/Rat/Dog/Chick/Fugu/Zfish Multiz Alignments & Conservation
Conservation
chimp
dog
mouse
rat
chicken
fugu
zebrafish
Fugu Blat
Fugu (Aug. 2002/fr1) Translated Blat Alignments
Simple Nucleotide Polymorphisms (SNPs)
Variation Track
Display a menu



Extended DNA Case/Color

Extended DNA Case/Color Options

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

Position chr8:39,625,889-39,909,460 Reverse complement ☐

Letters per line 60 Default case: ○ Upper ● Lower

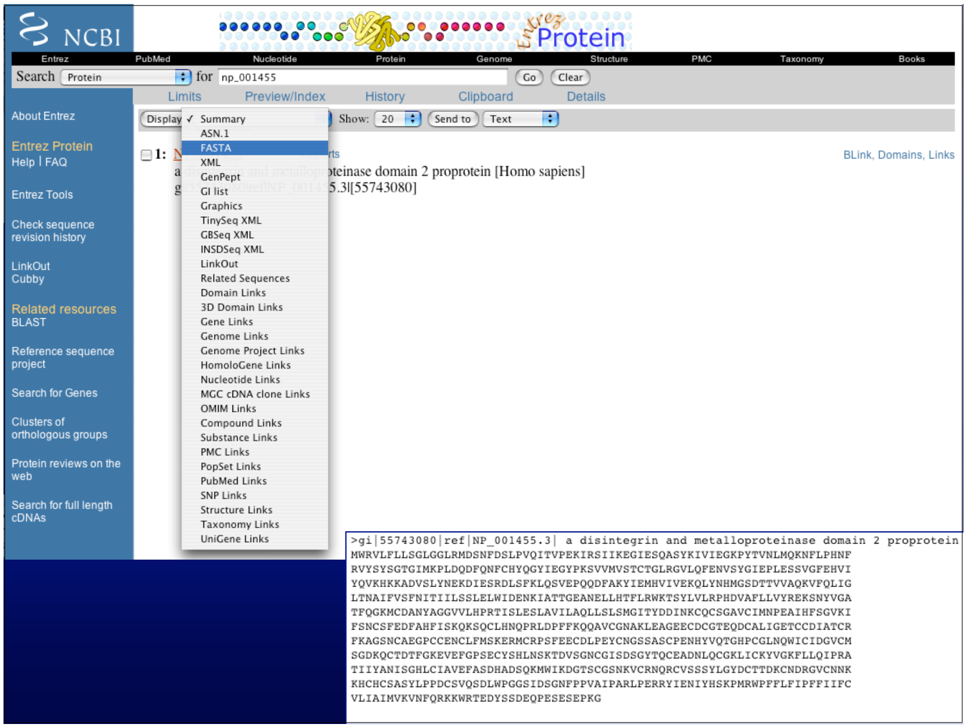
Track Name	Toggle Case	Under-line	Bold	Italic	Red	Green	Blue
ChromosomeBand	☐	☐	☐	☐	0	0	0
STS Markers	☐	☐	☐	☐	0	0	0
Known Genes	☐	☐	☐	☐	0	0	0
RefSeq Genes	☑	☐	☐	☐	255	0	0
Ensembl Genes	☐	☐	☐	☐	0	0	0
Genscan Genes	☐	☐	☐	☐	0	0	0
Human mRNAs	☐	☐	☐	☐	0	0	0
Spliced ESTs	☐	☐	☐	☐	0	0	0
Conservation	☐	☐	☐	☐	0	0	0
Fugu Blat	☐	☐	☐	☐	0	0	0
SNPs	☐	☑	☐	☐	0	255	0
RepeatMasker	☐	☐	☐	☐	0	0	0

SNPs in RefSeq exon

```
tgttactaattttgctgacctcatgaaaggaatcagcatgtgtaggactgaaataatat
catcctttactttaaatgtttatatacttgaaaatgtatctttttttcagcagcatatggt
gaaaaattttccaaatttttaggatcattgcctttgattgatttcagaccatcacactat
aacataacacataccatttgcaatttGAAATGTAACAGTGGCTTTTAATCATAGTACCACC
TCATTACATTATGATATTTCCTTATACCATGCCCCTACTGACAGAATTAGTGTCTACTAA
CCAACAGAAATAGCCTAAGATAACTTTAAAAAAACAGAAATTTGTAATGTAAAACTCATT
ATCATGCATTTTTAAATAAGTAAATTATCTGTGAAATTATACATATATATCATATTTTCT
GGTCAAATTAAAATCACAATATGCAGGTTTGAATCCCAGGAAACCAGGTAATGTACATTC
TTTTCTTTCCATTTTTTTGTGTCCATCCATCACAGATAATTACTCTAGAAGActgaataa
atgaaaaaaaaagtcaccgttatcaacaagttgaaatgatagctagttcaacaaatagca
tgcaagtaaccttattgcaaagtagacttatatcatgcaatttattttagccaacattt
taaatatacaaatgctaaagaagtaaaattaactaaagaaaataattaaaaaataacact
gtaataaattttaagtggaaaatttcaaaatcatttataacttgatcatgtctattata
aattctgaaaaatataattccgaaaaatagtctccaatttattttattatatatagcta
tatgactgtcactacttaatatatatttatttaatatattacgttctttctctctcctcc
ccccaacccccccatccttaaaatatttggaactcactcatgtcctttgcagggacatgg
atgaagctggaagccatcattctcaacaaacacacacaggaacataaaaccaaacaccgc
atgttctcactcataagtgggagctgaatcatgaaaacacatggacacagggaggggaac
atcacacaccagaacctgtcggggttgggaggcaaggggaggaagagcattgagaccaa
tacctaatgcatgtggggcttaaaacctagatgacgggttgataggtgcagcaaaccacc
atggcacatgtatacctatgtaacaaacctgcacattctgcacctgtgtcccagtactta
aagtaaaataaattttaaaaaaatgaaagaaactcacAGTGATATCATGGCATCTCTGTT
GTCCAGACTACCCTTTAGGTTCACTCTCACTTTCAGGTTGCctgcatattaaaaataaaa
ataatattatactttcttcaagatgattttacaaacctctgtattatctgtcagtcttga
aacagttctgaattttacctaagaactgggttcacccattatttgagaataccctatca
ctgagaacatctctgccctggtaatcaagactaataatgaacatctttgcatcagtgaat
atgaaaaaaaagcacagtgatatatgtttacccgaaaagaaacttagatcattgttatat
tcatagtattattggttaagtatctttaaactttttgccattcattttcaaccctattaa
tttctattcattttactagctatgctttttttccctgaagatatttgatatcagttattcc
acaatagttatccaatataagaaaaccactttagtatatttaaataatgactaagttatt
actcgtttaggtgctgtatattatttcagcaatttcaaatcaacttttaaaatatattcc
attttcatttacattgtcattttaagtattttaatgaaattcaattcatttcaagtaaaa
tgaaggacatagttatataactcccagggagaaatcatttggtagcattttccttgaaaa
taaataaaccagttttaatgtttacttgtattttatagcacatattcagatatttttat
ttttaatagaagttatttgttttatattttctgtagactagctgtggctatcatctagat
```

UCSC

Find a chicken homolog of a human protein

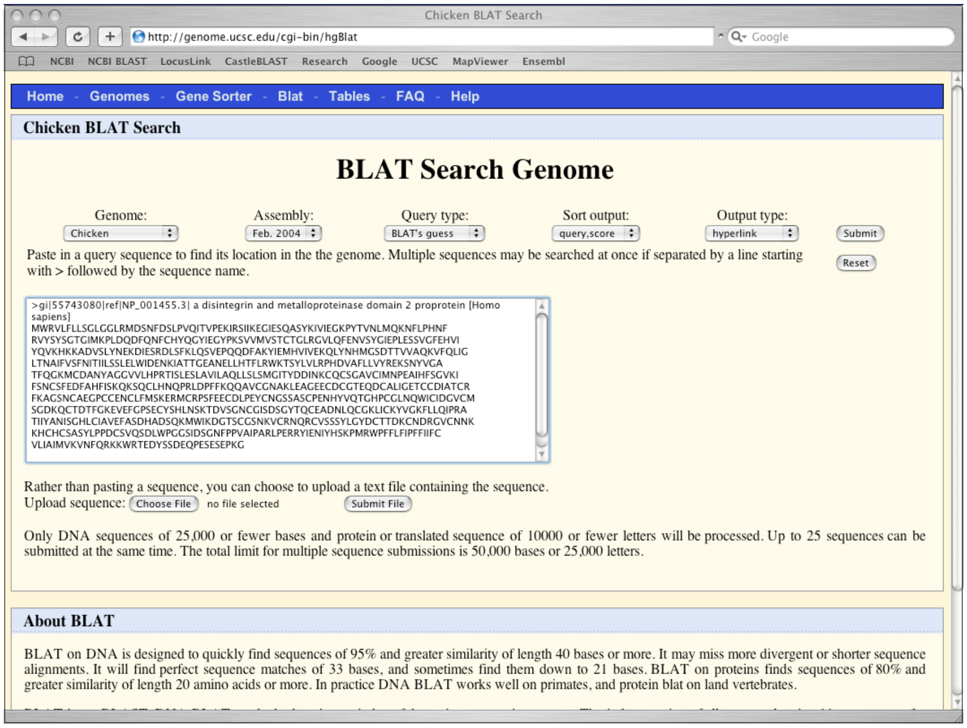


Search Protein for np_001455

Show: 20 Text

☐ 1: N… a …teinase domain 2 proprotein [Homo sapiens]
g…5.3|[55743080]

```
>gi|55743080|ref|NP_001455.3| a disintegrin and metalloproteinase domain 2 proprotein
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKIVIEGKPYTVNLMQKNFLPHNF
RVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPLESSVGFEHVI
YQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDFAKYIEMHVIVEKQLYNHMGSDTTVVAQKVFQLIG
LTNAIFVSFNITIILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLRPHDVAFLLVYREKSNYVGA
TFQGKMCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAIHFSGVKI
FSNCSFEDFAHFISKQKSQCLHNQPRLDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCM
SGDKQCTDTFGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYTQCEADNLQCGKLICKYVGKFLLQIPRA
TIIYANISGHLCIAVEFASDHADSQKMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTDKCNDRGVCNNK
KHCHCSASYLPPDCSVQSDLWPGGSIDSGNFPPVAIPARLPERRYIENIYHSKPMRWPFFLFIPFFIIFC
VLIAIMVKVNFQRKKWRTEDYSSDEQPESESEPKG
```



Chicken BLAT Search

BLAT Search Genome

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

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

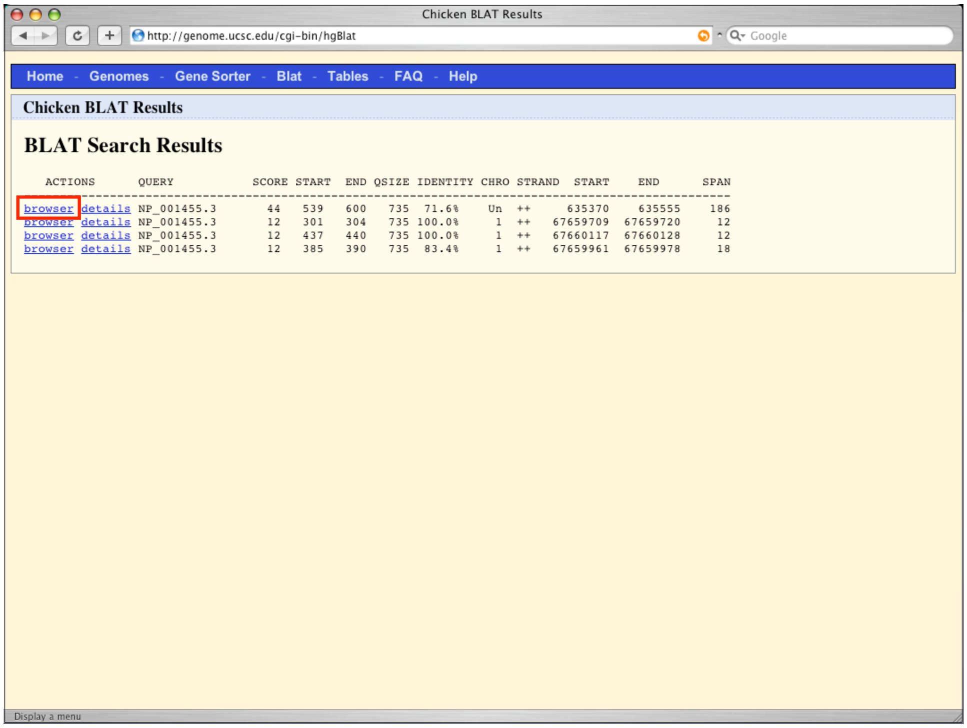
```
>gi|55743080|ref|NP_001455.3| a disintegrin and metalloproteinase domain 2 proprotein [Homo
sapiens]
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKIVIEGKPYTVNLMQKNFLPHNF
RVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPLESSVGFEHVI
YQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDFAKYIEMHVIVEKQLYNHMGSDTTVVAQKVFQLIG
LTNAIFVSFNITIILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLRPHDVAFLLVYREKSNYVGA
TFQGKMCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAIHFSGVKI
FSNCSFEDFAHFISKQKSQCLHNQPRLDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCM
SGDKQCTDTFGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYTQCEADNLQCGKLICKYVGKFLLQIPRA
TIIYANISGHLCIAVEFASDHADSQKMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTDKCNDRGVCNNK
KHCHCSASYLPPDCSVQSDLWPGGSIDSGNFPPVAIPARLPERRYIENIYHSKPMRWPFFLFIPFFIIFC
VLIAIMVKVNFQRKKWRTEDYSSDEQPESESEPKG
```

Rather than pasting a sequence, you can choose to upload a text file containing the sequence.
Upload sequence: no file selected

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.

About BLAT

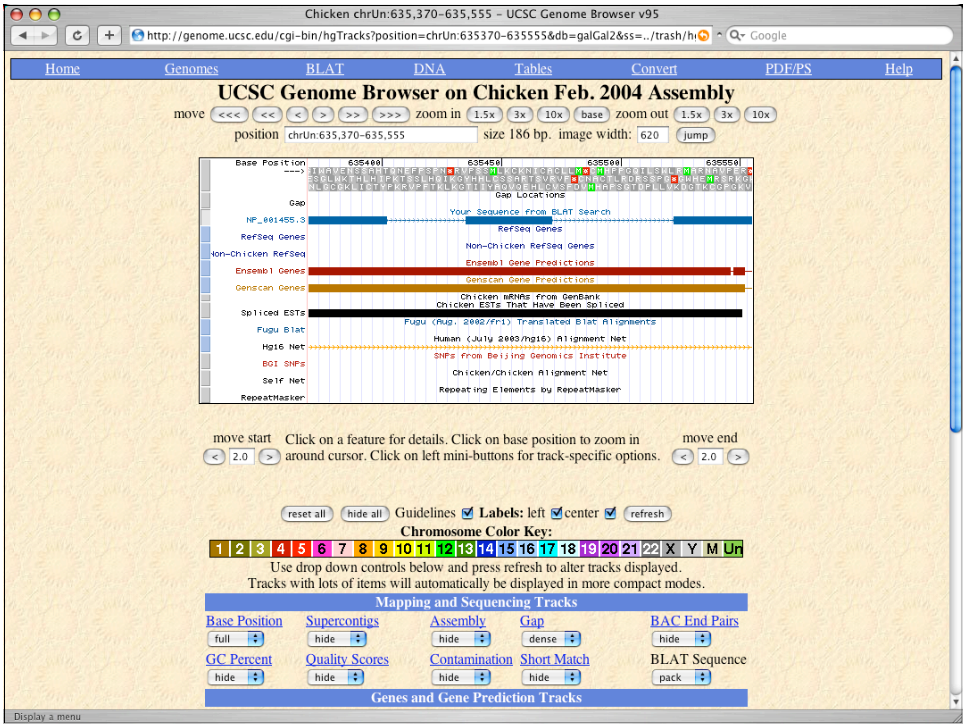
BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 40 bases or more. It may miss more divergent or shorter sequence alignments. It will find perfect sequence matches of 33 bases, and sometimes find them down to 21 bases. BLAT on proteins finds sequences of 80% and 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

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

position chrUn:635,370-635,555 size 186 bp. image width: 620

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

Guidelines ☑ **Labels:** left ☑ center ☑

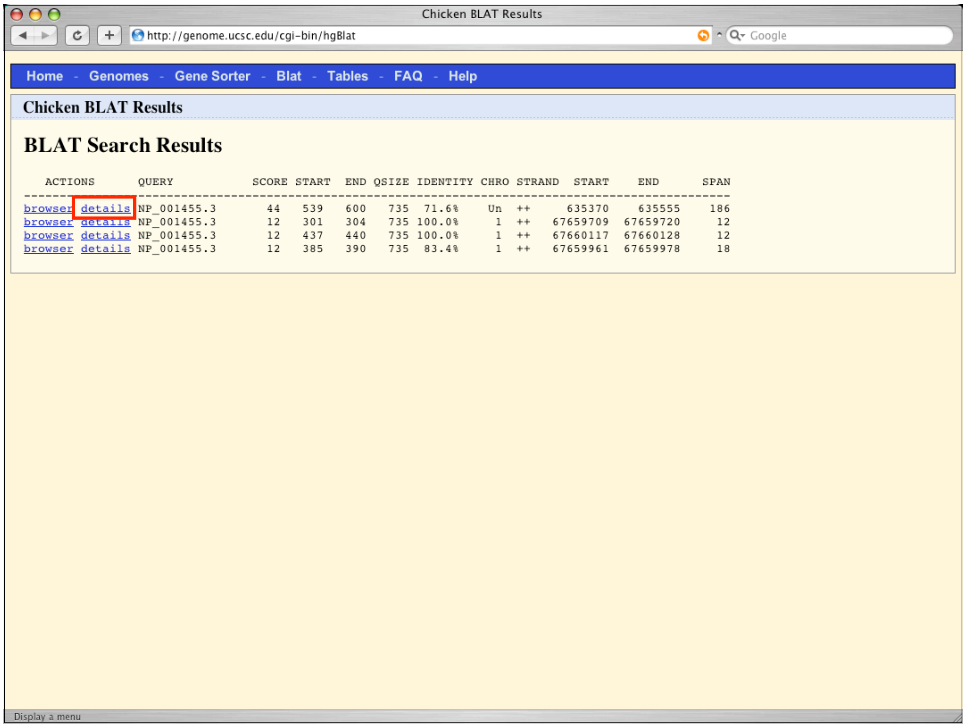
Chromosome Color Key:

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

Mapping and Sequencing Tracks

Base Position	Supercontigs	Assembly	Gap	BAC End Pairs
full	hide	hide	dense	hide
GC Percent	Quality Scores	Contamination	Short Match	BLAT Sequence
hide	hide	hide	hide	pack

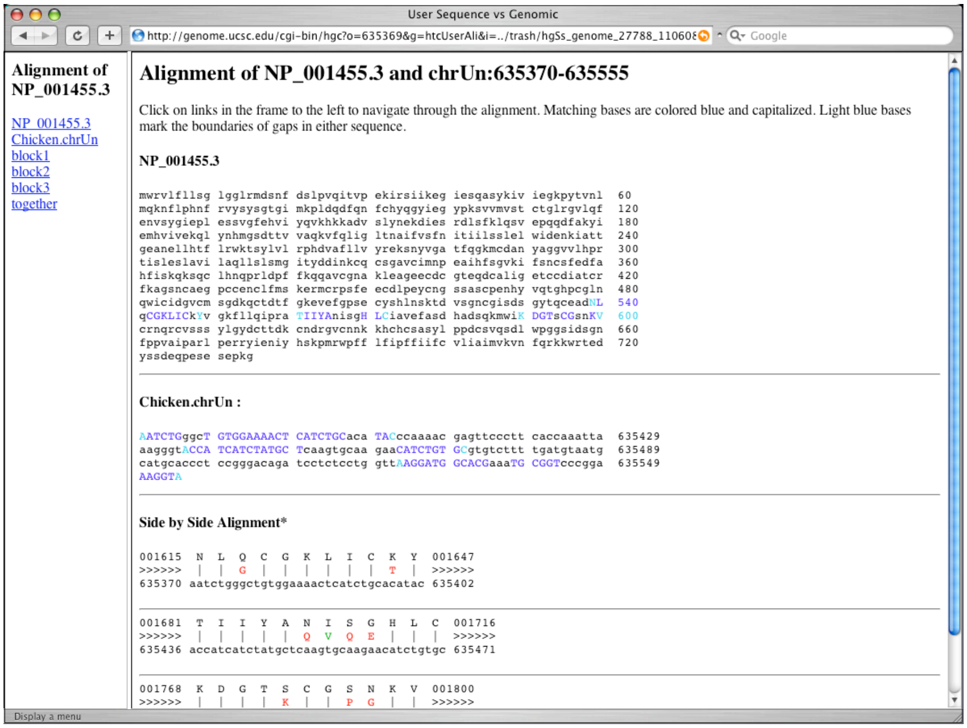
Genes and Gene Prediction Tracks



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 and chrUn:635370-635555

Click on links in the frame to the left to navigate through the alignment. Matching bases are colored blue and capitalized. Light blue bases mark the boundaries of gaps in either sequence.

NP_001455.3

```
mwrvlfllsg lgglrmdsnf dslpvqitvp ekirsiikeg iesqasykiv iegkpytvnl  60
mqknflphnf rvysysgtgi mkpldqdfqn fchyqgyieg ypksvvmvst ctglrgvlqf  120
envsygiepl essvgfehvi yqvkhkkadv slynekdies rdlsfklqsv epqqdfakyi  180
emhvivekql ynhmgsdttv vaqkvfqlig ltnaifvsfn itiilsslel widenkiatt  240
geanellhtf lrwktsylvl rphdvafllv yreksnyvga tfqgkmcdan yaggvvlhpr  300
tisleslavi laqllslsmg ityddinkcq csgavcimnp eaihfsgvki fsncsfedfa  360
hfiskqksqc lhnqprldpf fkqqavcgna kleageecdc gteqdcalig etccdiatcr  420
fkagsncaeg pccenclfms kermcrpsfe ecdlpeycng ssascpenhy vqtghpcgln  480
qwicidgvcm sgdkqctdtf gkevefgpse cyshlnsktd vsgncgisds gytqceadNL  540
qCGKLICkYv gkfllqipra TIIYAnisgH LCiavefasd hadsqkmwiK DGTsCGsnKV  600
crnqrcvsss ylgydcttdk cndrgvcnnk khchcsasyl ppdcsvqsdl wpggsidsgn  660
fppvaiparl perryieniy hskpmrwpff lfipffiifc vliaimvkvn fqrkkwrted  720
yssdeqpese sepkg
```

Chicken.chrUn :

```
AATCTGggcT GTGGAAAACT CATCTGCaca TACccaaaac gagttccctt caccaaatta  635429
aagggtACCA TCATCTATGC Tcaagtgcaa gaaCATCTGT GCgtgtcttt tgatgtaatg  635489
catgcaccct ccgggacaga tcctctcctg gttAAGGATG GCACGaaaTG CGGTcccgga  635549
AAGGTA
```

Side by Side Alignment*

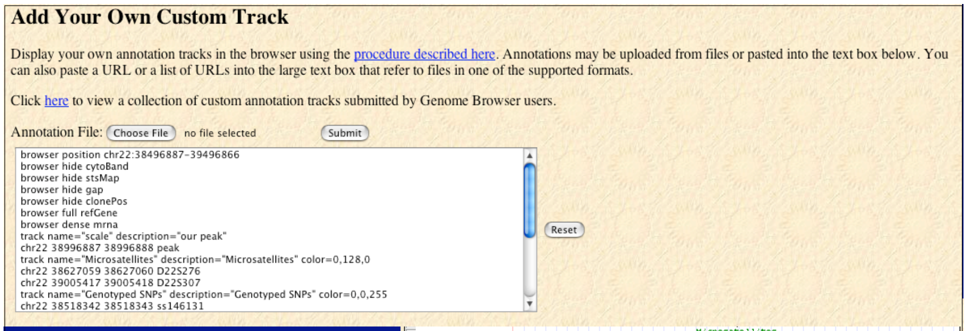
```
001615  N  L  Q  C  G  K  L  I  C  K  Y  001647
>>>>>>  |  |  G  |  |  |  |  |  |  T  |  >>>>>>
635370 aatctgggctgtggaaaactcatctgcacatac 635402
```

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

```
001768  K  D  G  T  S  C  G  S  N  K  V  001800
>>>>>>  |  |  |  |  K  |  |  P  G  |  |  >>>>>>
```

UCSC

Add your own custom tracks



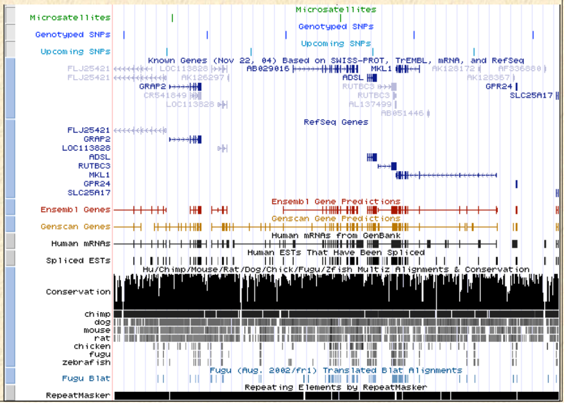
Add Your Own Custom Track

Display your own annotation tracks in the browser using the procedure described here. Annotations may be uploaded from files or pasted into the text box below. You can also paste a URL or a list of URLs into the large text box that refer to files in one of the supported formats.

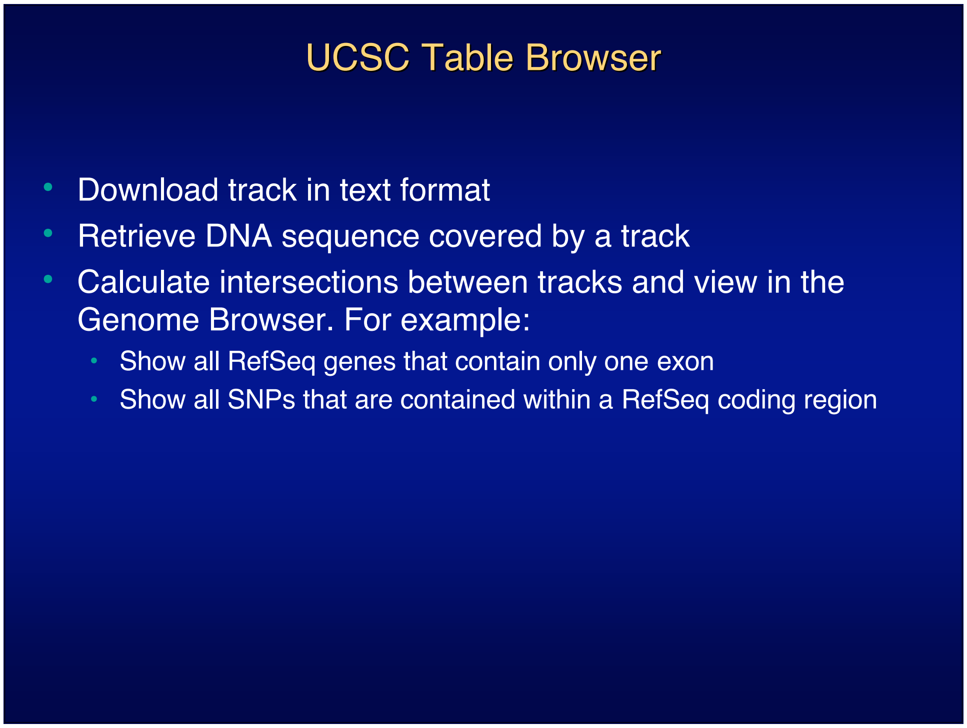
Click here to view a collection of custom annotation tracks submitted by Genome Browser users.

Annotation File:

```
browser position chr22:38496887-39496866
browser hide cytoBand
browser hide stsMap
browser hide gap
browser hide clonePos
browser full refGene
browser dense mrna
track name="scale" description="our peak"
chr22 38996887 38996888 peak
track name="Microsatellites" description="Microsatellites" color=0,128,0
chr22 38627059 38627060 D22S276
chr22 39005417 39005418 D22S307
track name="Genotyped SNPs" description="Genotyped SNPs" color=0,0,255
chr22 38518342 38518343 ss146131
```

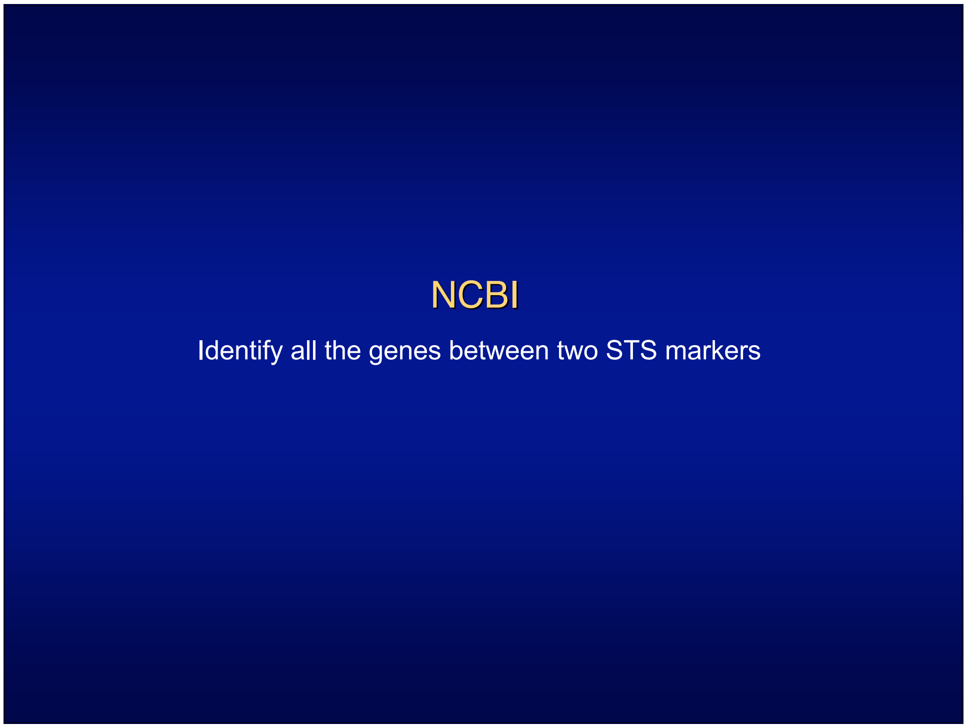


Nature Genetics User's Guide, Question 7



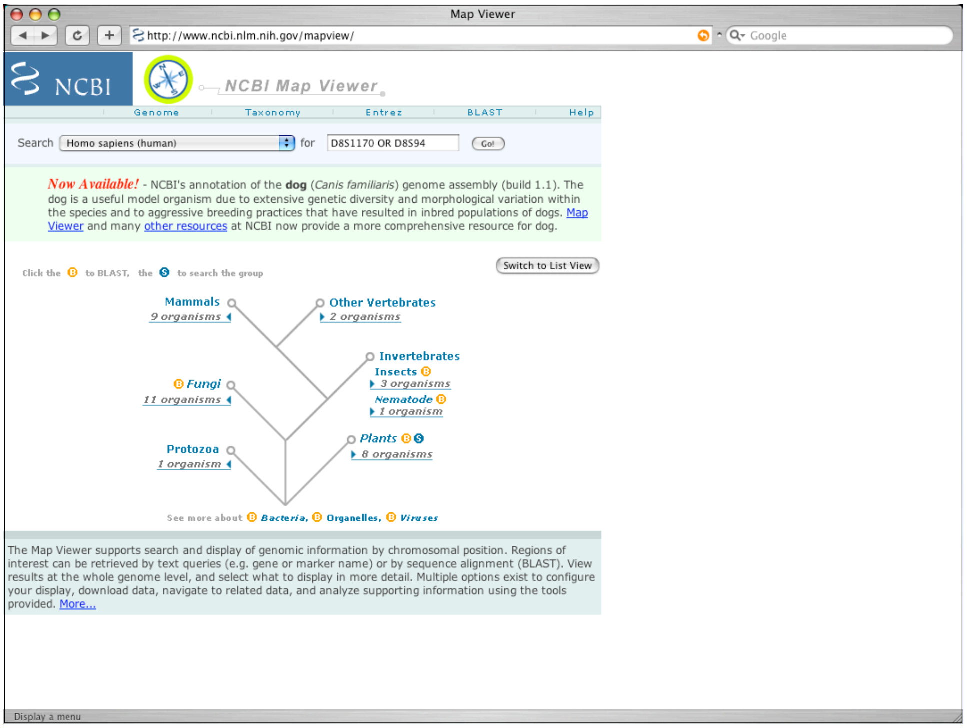
UCSC Table Browser

- 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 all SNPs that are contained within a RefSeq coding region



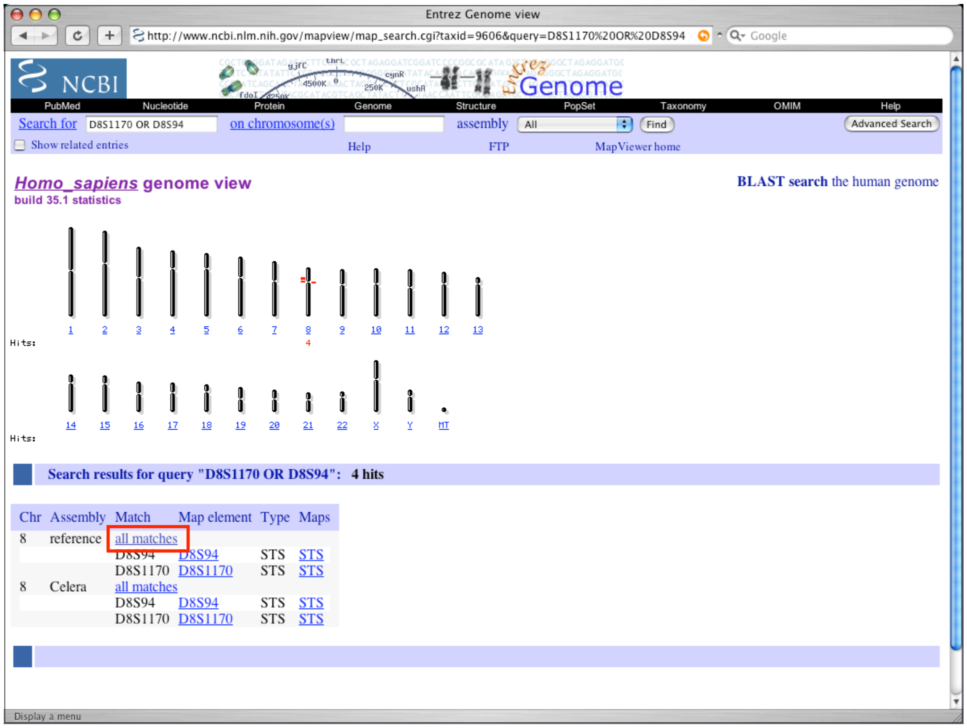
NCBI

Identify all the genes between two STS markers



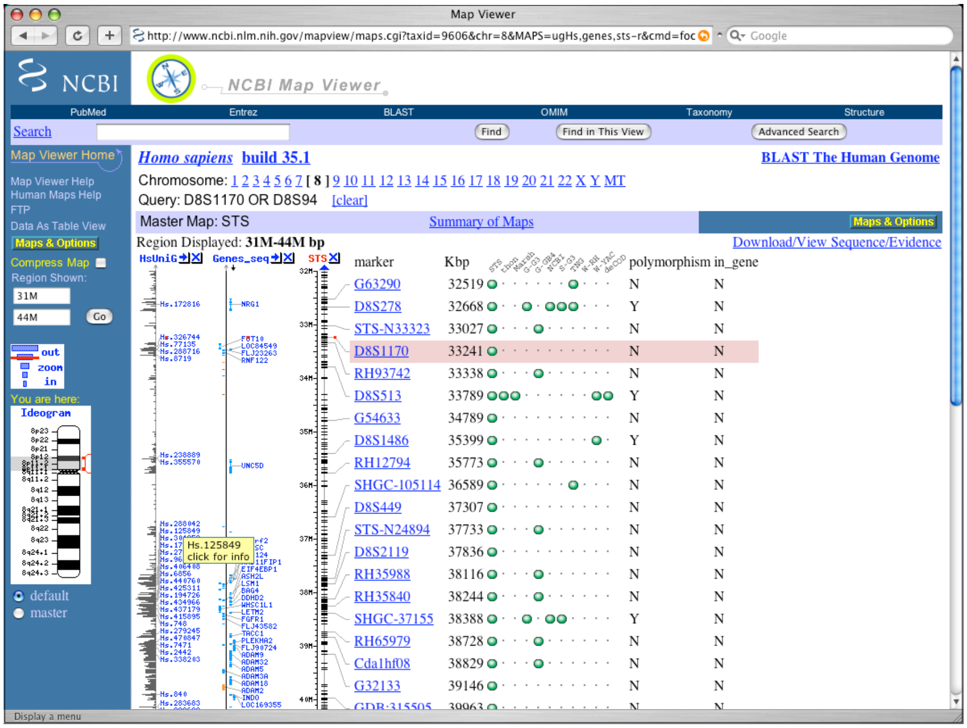
Now Available! - NCBI's annotation of the **dog** (*Canis familiaris*) genome assembly (build 1.1). The dog is a useful model organism due to extensive genetic diversity and morphological variation within the species and to aggressive breeding practices that have resulted in inbred populations of dogs. Map Viewer and many other resources at NCBI now provide a more comprehensive resource for dog.

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



Search results for query "D8S1170 OR D8S94": 4 hits

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



Homo sapiens build 35.1 — BLAST The Human Genome

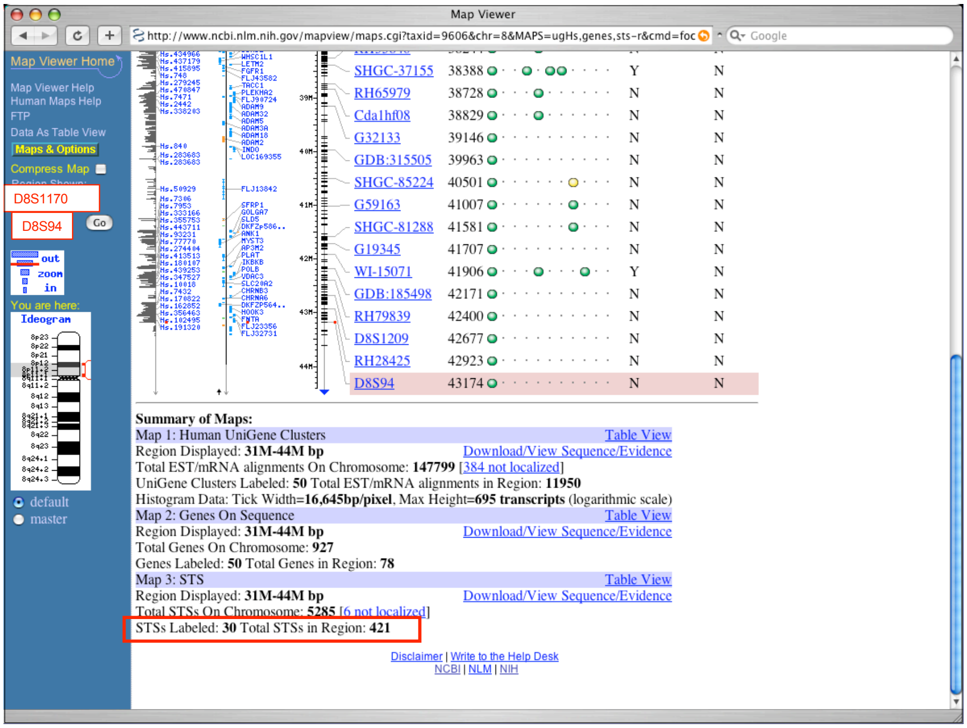
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT

Query: D8S1170 OR D8S94 [clear]

Master Map: STS — Summary of Maps — Maps & Options

Region Displayed: **31M-44M bp** — Download/View Sequence/Evidence

marker	Kbp	STS	thon	Marsh	G-G3	G-GB4	NCBI	S-G3	TNG	W-RH	W-YAC	deCOD	polymorphism	in_gene
G63290	32519	●	·	·	·	·	·	·	●	·	·	·	N	N
D8S278	32668	●	·	·	●	·	●	●	●	·	·	·	Y	N
STS-N33323	33027	●	·	·	·	●	·	·	·	·	·	·	N	N
D8S1170	33241	●	·	·	·	·	·	·	·	·	·	·	N	N
RH93742	33338	●	·	·	·	●	·	·	·	·	·	·	N	N
D8S513	33789	●	●	●	·	·	·	·	·	·	●	●	Y	N
G54633	34789	●	·	·	·	·	·	·	·	·	·	·	N	N
D8S1486	35399	●	·	·	·	·	·	·	·	·	●	·	Y	N
RH12794	35773	●	·	·	·	●	·	·	·	·	·	·	N	N
SHGC-105114	36589	●	·	·	·	·	·	·	●	·	·	·	N	N
D8S449	37307	●	·	·	·	·	·	·	·	·	·	·	N	N
STS-N24894	37733	●	·	·	·	●	·	·	·	·	·	·	N	N
D8S2119	37836	●	·	·	·	·	·	·	·	·	·	·	N	N
RH35988	38116	●	·	·	·	●	·	·	·	·	·	·	N	N
RH35840	38244	●	·	·	·	●	·	·	·	·	·	·	N	N
SHGC-37155	38388	●	·	·	●	·	●	●	·	·	·	·	Y	N
RH65979	38728	●	·	·	·	●	·	·	·	·	·	·	N	N
Cda1hf08	38829	●	·	·	·	●	·	·	·	·	·	·	N	N
G32133	39146	●	·	·	·	·	·	·	·	·	·	·	N	N
GDB:315505	39963	●	·	·	·	·	·	·	·	·	·	·	N	N



marker	Kbp	STS	thon	Marsh	G-G3	G-GB4	NCBI	S-G3	TNG	W-RH	W-YAC	deCOD	polymorphism	in_gene
SHGC-37155	38388	●	·	·	●	·	●	●	·	·	·	·	Y	N
RH65979	38728	●	·	·	·	●	·	·	·	·	·	·	N	N
Cda1hf08	38829	●	·	·	·	●	·	·	·	·	·	·	N	N
G32133	39146	●	·	·	·	·	·	·	·	·	·	·	N	N
GDB:315505	39963	●	·	·	·	·	·	·	·	·	·	·	N	N
SHGC-85224	40501	●	·	·	·	·	·	·	🟡	·	·	·	N	N
G59163	41007	●	·	·	·	·	·	·	●	·	·	·	N	N
SHGC-81288	41581	●	·	·	·	·	·	·	●	·	·	·	N	N
G19345	41707	●	·	·	·	·	·	·	·	·	·	·	N	N
WI-15071	41906	●	·	·	·	●	·	·	·	●	·	·	Y	N
GDB:185498	42171	●	·	·	·	·	·	·	·	·	·	·	N	N
RH79839	42400	●	·	·	·	·	·	·	·	·	·	·	N	N
D8S1209	42677	●	·	·	·	·	·	·	·	·	·	·	N	N
RH28425	42923	●	·	·	·	·	·	·	·	·	·	·	N	N
D8S94	43174	●	·	·	·	·	·	·	·	·	·	·	N	N

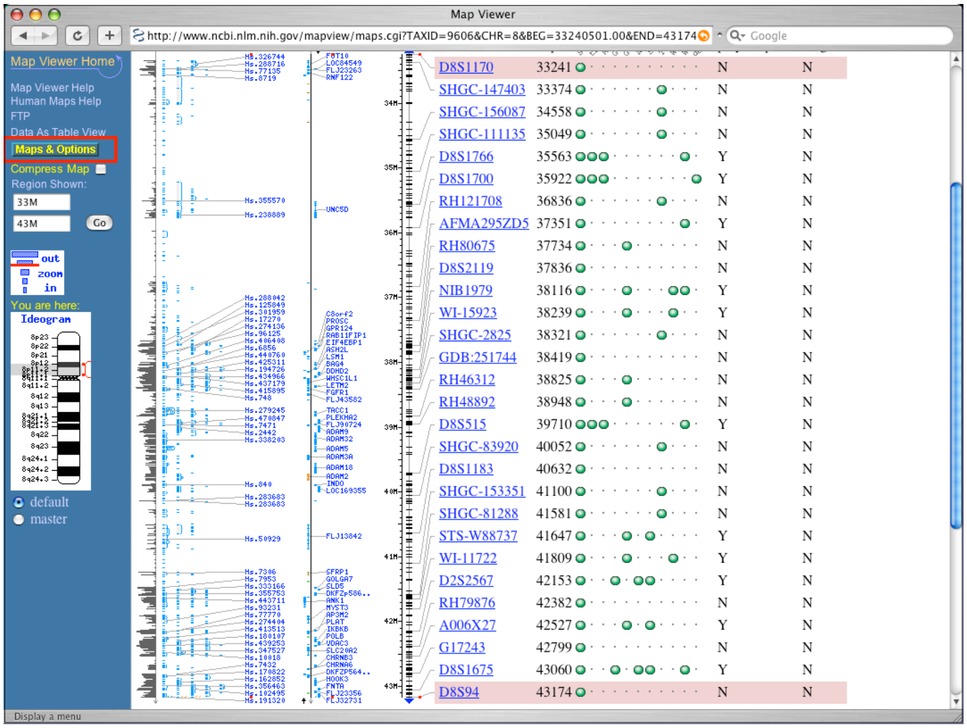
Summary of Maps:

Map 1: Human UniGene Clusters — Table View
Region Displayed: **31M-44M bp** — Download/View Sequence/Evidence
Total EST/mRNA alignments On Chromosome: **147799** [384 not localized]
UniGene Clusters Labeled: **50** Total EST/mRNA alignments in Region: **11950**
Histogram Data: Tick Width=**16,645bp/pixel**, Max Height=**695 transcripts** (logarithmic scale)

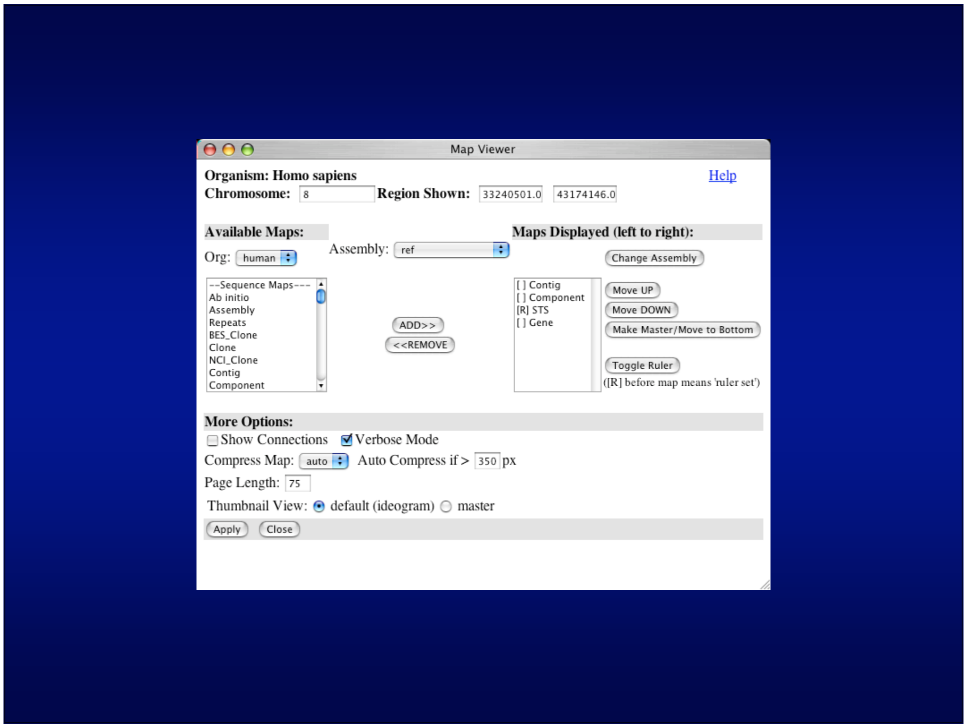
Map 2: Genes On Sequence — Table View
Region Displayed: **31M-44M bp** — Download/View Sequence/Evidence
Total Genes On Chromosome: **927**
Genes Labeled: **50** Total Genes in Region: **78**

Map 3: STS — Table View
Region Displayed: **31M-44M bp** — Download/View Sequence/Evidence
Total STSs On Chromosome: **5285** [6 not localized]
STSs Labeled: **30** Total STSs in Region: **421**

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



D8S1170	33241	●··········	N	N
SHGC-147403	33374	●······●···	N	N
SHGC-156087	34558	●······●···	N	N
SHGC-111135	35049	●······●···	N	N
D8S1766	35563	●●●······●·	Y	N
D8S1700	35922	●●●·······●	Y	N
RH121708	36836	●······●···	N	N
AFMA295ZD5	37351	●········●·	Y	N
RH80675	37734	●···●······	N	N
D8S2119	37836	●··········	N	N
NIB1979	38116	●···●···●●·	Y	N
WI-15923	38239	●···●···●··	Y	N
SHGC-2825	38321	●······●···	N	N
GDB:251744	38419	●··········	N	N
RH46312	38825	●···●······	N	N
RH48892	38948	●···●······	N	N
D8S515	39710	●●●······●·	Y	N
SHGC-83920	40052	●······●···	N	N
D8S1183	40632	●··········	N	N
SHGC-153351	41100	●······●···	N	N
SHGC-81288	41581	●······●···	N	N
STS-W88737	41647	●···●·●····	Y	N
WI-11722	41809	●···●···●··	Y	N
D2S2567	42153	●··●·●●····	Y	N
RH79876	42382	●··········	N	N
A006X27	42527	●···●·●····	Y	N
G17243	42799	●··········	N	N
D8S1675	43060	●··●·●●··●·	Y	N
D8S94	43174	●··········	N	N



Organism: Homo sapiens
Chromosome: 8 **Region Shown:** 33240501.0 43174146.0

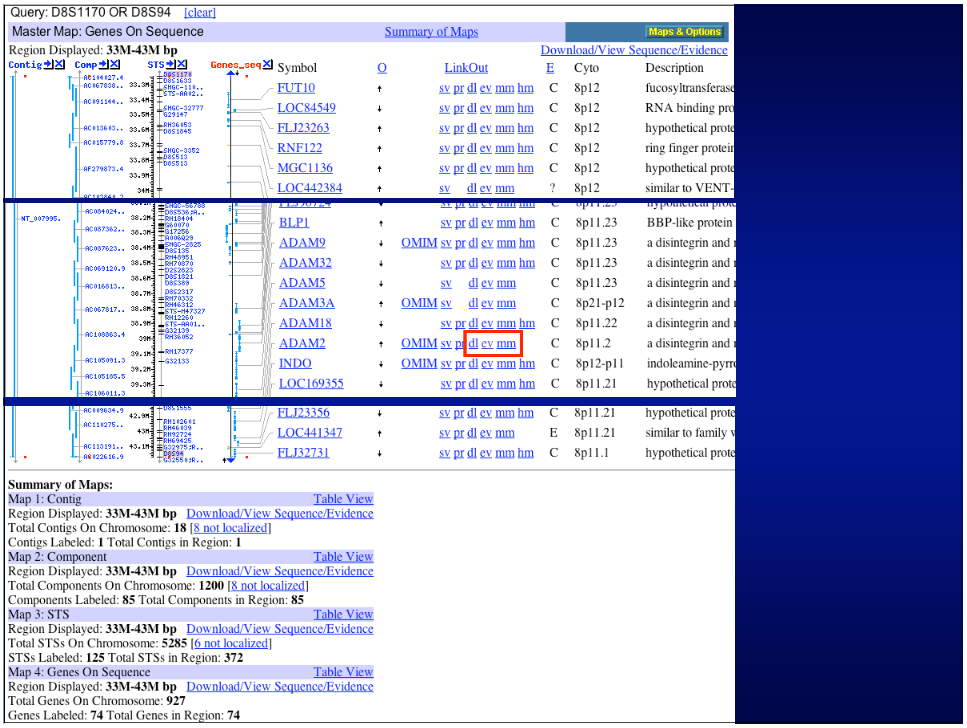
Available Maps:
Org: human
Assembly: ref

--Sequence Maps---
Ab initio
Assembly
Repeats
BES_Clone
Clone
NCI_Clone
Contig
Component

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

([R] before map means 'ruler set')

More Options:
☐ Show Connections ☑ Verbose Mode
Compress Map: auto Auto Compress if > 350 px
Page Length: 75
Thumbnail View: ● default (ideogram) ○ master



Query: D8S1170 OR D8S94 [clear]

Master Map: Genes On Sequence — Summary of Maps — Maps & Options

Region Displayed: **33M-43M bp** — Download/View Sequence/Evidence

Symbol	O	LinkOut	E	Cyto	Description
FUT10	↑	sv pr dl ev mm hm	C	8p12	fucosyltransferase
LOC84549	↓	sv pr dl ev mm hm	C	8p12	RNA binding pro
FLJ23263	↑	sv pr dl ev mm hm	C	8p12	hypothetical prote
RNF122	↑	sv pr dl ev mm hm	C	8p12	ring finger protein
MGC1136	↑	sv pr dl ev mm hm	C	8p12	hypothetical prote
LOC442384	↑	sv dl ev mm	?	8p12	similar to VENT-
BLP1	↑	sv pr dl ev mm hm	C	8p11.23	BBP-like protein
ADAM9	↓	OMIM sv pr dl ev mm hm	C	8p11.23	a disintegrin and
ADAM32	↓	sv pr dl ev mm hm	C	8p11.23	a disintegrin and
ADAM5	↓	sv dl ev mm	C	8p11.23	a disintegrin and
ADAM3A	↑	OMIM sv dl ev mm	C	8p21-p12	a disintegrin and
ADAM18	↓	sv pr dl ev mm hm	C	8p11.22	a disintegrin and
ADAM2	↑	OMIM sv pr dl ev mm	C	8p11.2	a disintegrin and
INDO	↓	OMIM sv pr dl ev mm hm	C	8p12-p11	indoleamine-pyrr
LOC169355	↓	sv pr dl ev mm hm	C	8p11.21	hypothetical prote
FLJ23356	↓	sv pr dl ev mm hm	C	8p11.21	hypothetical prote
LOC441347	↑	sv pr dl ev mm	E	8p11.21	similar to family w
FLJ32731	↓	sv pr dl ev mm hm	C	8p11.1	hypothetical prote

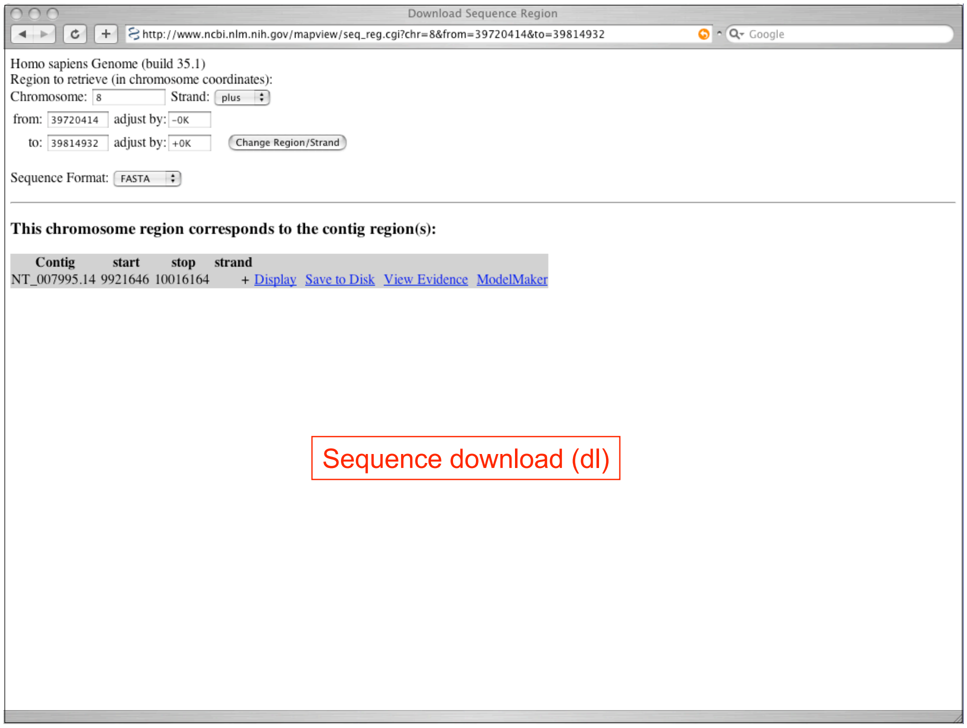
Summary of Maps:

Map 1: Contig — Table View
Region Displayed: **33M-43M bp** Download/View Sequence/Evidence
Total Contigs On Chromosome: **18** [8 not localized]
Contigs Labeled: **1** Total Contigs in Region: **1**

Map 2: Component — Table View
Region Displayed: **33M-43M bp** Download/View Sequence/Evidence
Total Components On Chromosome: **1200** [8 not localized]
Components Labeled: **85** Total Components in Region: **85**

Map 3: STS — Table View
Region Displayed: **33M-43M bp** Download/View Sequence/Evidence
Total STSs On Chromosome: **5285** [6 not localized]
STSs Labeled: **125** Total STSs in Region: **372**

Map 4: Genes On Sequence — Table View
Region Displayed: **33M-43M bp** Download/View Sequence/Evidence
Total Genes On Chromosome: **927**
Genes Labeled: **74** Total Genes in Region: **74**



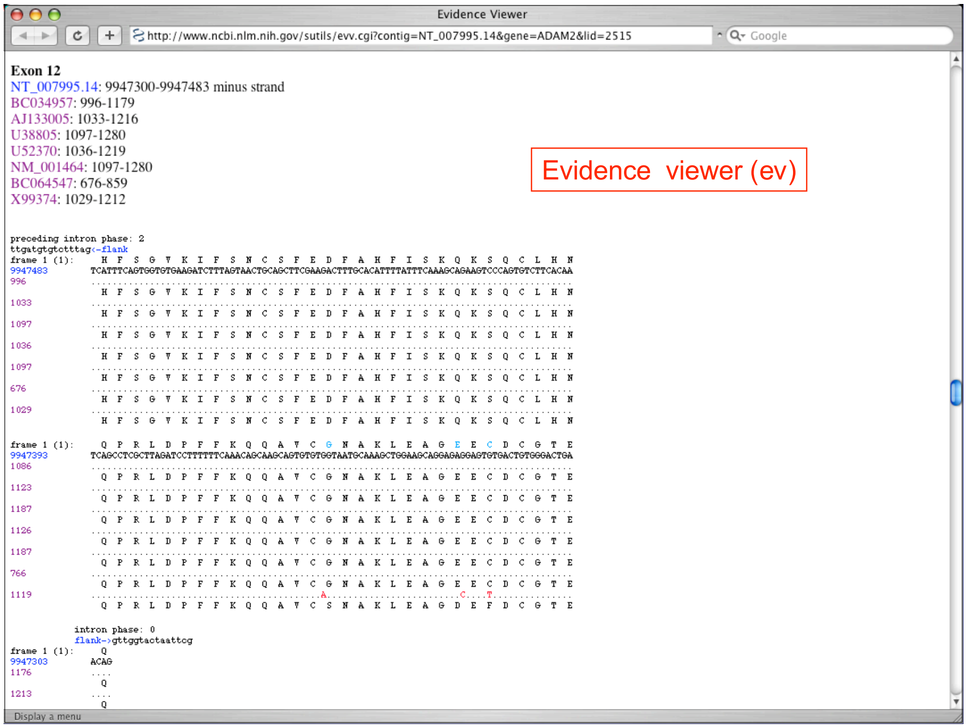
Homo sapiens Genome (build 35.1)
Region to retrieve (in chromosome coordinates):
Chromosome: 8 Strand: plus
from: 39720414 adjust by: -0K
to: 39814932 adjust by: +0K
Change Region/Strand

Sequence Format: FASTA

This chromosome region corresponds to the contig region(s):

Contig	start	stop	strand	
NT_007995.14	9921646	10016164	+	Display Save to Disk View Evidence ModelMaker

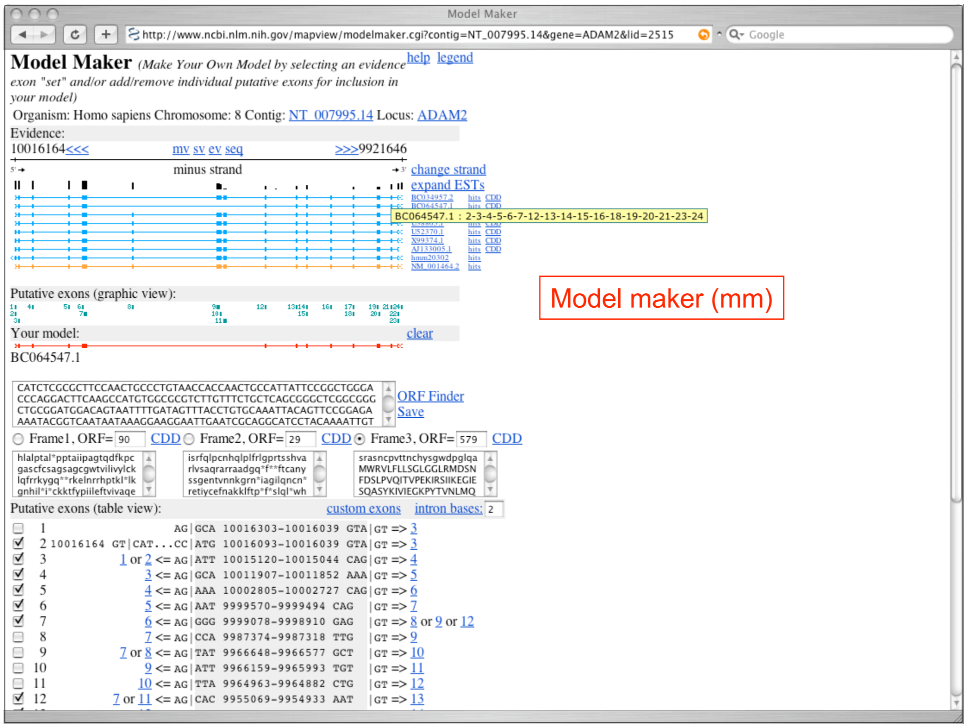
Sequence download (dl)



Exon 12
NT_007995.14: 9947300-9947483 minus strand
BC034957: 996-1179
AJ133005: 1033-1216
U38805: 1097-1280
U52370: 1036-1219
NM_001464: 1097-1280
BC064547: 676-859
X99374: 1029-1212

Evidence viewer (ev)

```
preceding intron phase: 2
ttgatgtgtctttag<-flank
frame 1 (1):   H  F  S  G  V  K  I  F  S  N  C  S  F  E  D  F  A  H  F  I  S  K  Q  K  S  Q  C  L  H  N
9947483       TCATTTCAGTGGTGTGAAGATCTTTAGTAACTGCAGCTTCGAAGACTTTGCACATTTTATTTCAAAGCAGAAGTCCCAGTGTCTTCACAA
996           ....................................................................................
               H  F  S  G  V  K  I  F  S  N  C  S  F  E  D  F  A  H  F  I  S  K  Q  K  S  Q  C  L  H  N
1033          ....................................................................................
               H  F  S  G  V  K  I  F  S  N  C  S  F  E  D  F  A  H  F  I  S  K  Q  K  S  Q  C  L  H  N
1097          ....................................................................................
               H  F  S  G  V  K  I  F  S  N  C  S  F  E  D  F  A  H  F  I  S  K  Q  K  S  Q  C  L  H  N
1036          ....................................................................................
               H  F  S  G  V  K  I  F  S  N  C  S  F  E  D  F  A  H  F  I  S  K  Q  K  S  Q  C  L  H  N
1097          ....................................................................................
               H  F  S  G  V  K  I  F  S  N  C  S  F  E  D  F  A  H  F  I  S  K  Q  K  S  Q  C  L  H  N
676           ....................................................................................
               H  F  S  G  V  K  I  F  S  N  C  S  F  E  D  F  A  H  F  I  S  K  Q  K  S  Q  C  L  H  N
1029          ....................................................................................
               H  F  S  G  V  K  I  F  S  N  C  S  F  E  D  F  A  H  F  I  S  K  Q  K  S  Q  C  L  H  N

frame 1 (1):   Q  P  R  L  D  P  F  F  K  Q  Q  A  V  C  G  N  A  K  L  E  A  G  E  E  C  D  C  G  T  E
9947393       TCAGCCTCGCTTAGATCCTTTTTTCAAACAGCAAGCAGTGTGTGGTAATGCAAAGCTGGAAGCAGGAGAGGAGTGTGACTGTGGGACTGA
1086          ....................................................................................
               Q  P  R  L  D  P  F  F  K  Q  Q  A  V  C  G  N  A  K  L  E  A  G  E  E  C  D  C  G  T  E
1123          ....................................................................................
               Q  P  R  L  D  P  F  F  K  Q  Q  A  V  C  G  N  A  K  L  E  A  G  E  E  C  D  C  G  T  E
1187          ....................................................................................
               Q  P  R  L  D  P  F  F  K  Q  Q  A  V  C  G  N  A  K  L  E  A  G  E  E  C  D  C  G  T  E
1126          ....................................................................................
               Q  P  R  L  D  P  F  F  K  Q  Q  A  V  C  G  N  A  K  L  E  A  G  E  E  C  D  C  G  T  E
1187          ....................................................................................
               Q  P  R  L  D  P  F  F  K  Q  Q  A  V  C  G  N  A  K  L  E  A  G  E  E  C  D  C  G  T  E
766           ....................................................................................
               Q  P  R  L  D  P  F  F  K  Q  Q  A  V  C  G  N  A  K  L  E  A  G  E  E  C  D  C  G  T  E
1119          ........................................A.......................C....T..............
               Q  P  R  L  D  P  F  F  K  Q  Q  A  V  C  S  N  A  K  L  E  A  G  D  E  F  D  C  G  T  E

        intron phase: 0
        flank->gttggtactaattcg
frame 1 (1):   Q
9947303       ACAG
1176          ....
               Q
1213          ....
               Q
```



Model Maker *(Make Your Own Model by selecting an evidence exon "set" and/or add/remove individual putative exons for inclusion in your model)* help legend

Organism: Homo sapiens Chromosome: 8 Contig: NT_007995.14 Locus: ADAM2

Evidence:
10016164<<< mv sv ev seq >>>9921646
minus strand
change strand
expand ESTs

BC064547.1 : 2-3-4-5-6-7-12-13-14-15-16-18-19-20-21-23-24

Putative exons (graphic view):

Your model: clear

BC064547.1

```
CATCTCGCGCTTCCAACTGCCCTGTAACCACCAACTGCCATTATTCCGGCTGGGA
CCCAGGACTTCAAGCCATGTGGCGCGTCTTGTTTCTGCTCAGCGGGCTCGGCGGG
CTGCGGATGGACAGTAATTTTGATAGTTTACCTGTGCAAATTACAGTTCCGGAGA
AAATACGGTCAATAATAAAGGAAGGAATTGAATCGCAGGCATCCTACAAAATTGT
```

ORF Finder
Save

○ Frame1, ORF= 90 CDD ○ Frame2, ORF= 29 CDD ● Frame3, ORF= 579 CDD

```
hlalptal*pptaiipagtqdfkpc      isrfqlpcnhqlplfrlgprtsshva     srasncpvttnchysgwdpglqa
gascfcsagsagcgwtvilivylck      rlvsaqrarraadgq*f**ftcany      MWRVLFLLSGLGGLRMDSN
lqfrrkygq**rkelnrrhptkl*lk     ssgentvnnkgrn*iagilqncn*       FDSLPVQITVPEKIRSIIKEGIE
gnhil*i*ckktfypiileftvivaqe    retiycefnakklftp*f*slql*wh     SQASYKIVIEGKPYTVNLMQ
```

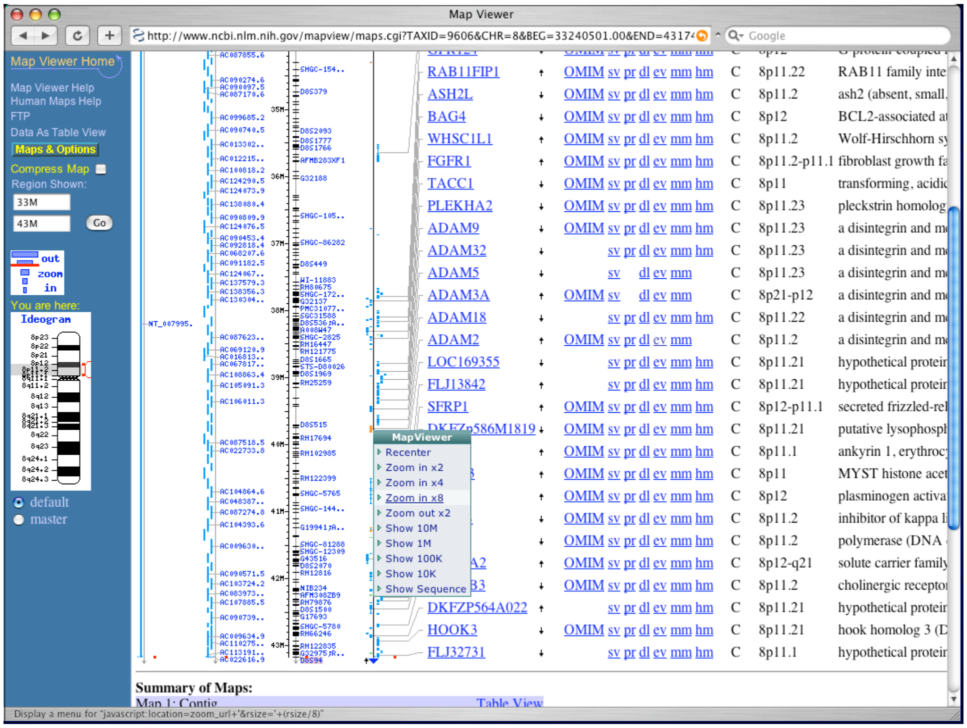
Putative exons (table view): custom exons intron bases: 2

```
☐  1                      AG|GCA 10016303-10016039 GTA|GT => 3
☑  2 10016164 GT|CAT...CC|ATG 10016093-10016039 GTA|GT => 3
☑  3             1 or 2 <= AG|ATT 10015120-10015044 CAG|GT => 4
☑  4                  3 <= AG|GCA 10011907-10011852 AAA|GT => 5
☑  5                  4 <= AG|AAA 10002805-10002727 CAG|GT => 6
☑  6                  5 <= AG|AAT 9999570-9999494 CAG  |GT => 7
☑  7                  6 <= AG|GGG 9999078-9998910 GAG  |GT => 8 or 9 or 12
☐  8                  7 <= AG|CCA 9987374-9987318 TTG  |GT => 9
☐  9             7 or 8 <= AG|TAT 9966648-9966577 GCT  |GT => 10
☐ 10                  9 <= AG|ATT 9966159-9965993 TGT  |GT => 11
☐ 11                 10 <= AG|TTA 9964963-9964882 CTG  |GT => 12
☑ 12            7 or 11 <= AG|CAC 9955069-9954933 AAT  |GT => 13
```

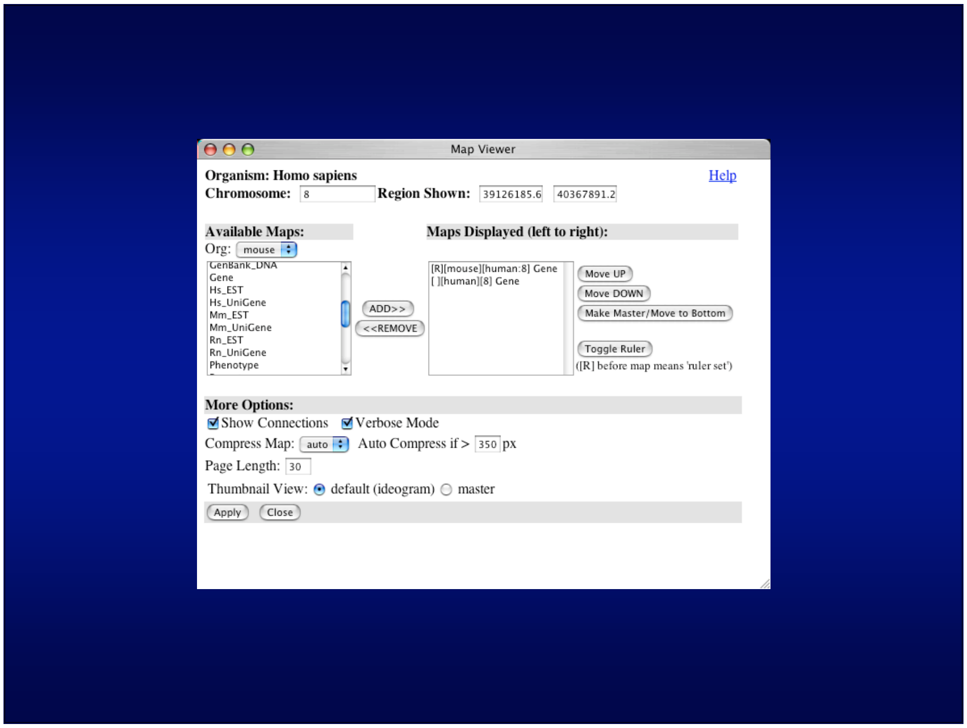
Model maker (mm)

NCBI

Change the maps displayed on the Map Viewer



RAB11FIP1	↑	OMIM sv pr dl ev mm hm	C	8p11.22	RAB11 family inte
ASH2L	↓	OMIM sv pr dl ev mm hm	C	8p11.2	ash2 (absent, small,
BAG4	↓	OMIM sv pr dl ev mm hm	C	8p12	BCL2-associated at
WHSC1L1	↑	OMIM sv pr dl ev mm hm	C	8p11.2	Wolf-Hirschhorn sy
FGFR1	↑	OMIM sv pr dl ev mm hm	C	8p11.2-p11.1	fibroblast growth fa
TACC1	↓	OMIM sv pr dl ev mm hm	C	8p11	transforming, acidic
PLEKHA2	↓	OMIM sv pr dl ev mm hm	C	8p11.23	pleckstrin homolog
ADAM9	↓	OMIM sv pr dl ev mm hm	C	8p11.23	a disintegrin and m
ADAM32	↓	sv pr dl ev mm hm	C	8p11.23	a disintegrin and m
ADAM5	↓	sv dl ev mm	C	8p11.23	a disintegrin and m
ADAM3A	↑	OMIM sv dl ev mm	C	8p21-p12	a disintegrin and m
ADAM18	↓	sv pr dl ev mm hm	C	8p11.22	a disintegrin and m
ADAM2	↑	OMIM sv pr dl ev mm	C	8p11.2	a disintegrin and m
LOC169355	↓	sv pr dl ev mm hm	C	8p11.21	hypothetical protein
FLJ13842	↑	sv pr dl ev mm hm	C	8p11.21	hypothetical protein
SFRP1	↑	OMIM sv pr dl ev mm hm	C	8p12-p11.1	secreted frizzled-rel
DKFZp586M1819	↓	OMIM sv pr dl ev mm hm	C	8p11.21	putative lysophosph
	↑	OMIM sv pr dl ev mm hm	C	8p11.1	ankyrin 1, erythroc
8	↑	OMIM sv pr dl ev mm hm	C	8p11	MYST histone acet
	↑	OMIM sv pr dl ev mm hm	C	8p12	plasminogen activa
	↓	OMIM sv pr dl ev mm hm	C	8p11.2	inhibitor of kappa l
	↓	OMIM sv pr dl ev mm hm	C	8p11.2	polymerase (DNA
A2	↑	OMIM sv pr dl ev mm hm	C	8p12-q21	solute carrier family
B3	↓	OMIM sv pr dl ev mm hm	C	8p11.2	cholinergic receptor
DKFZP564A022	↑	sv pr dl ev mm hm	C	8p11.21	hypothetical protein
HOOK3	↓	OMIM sv pr dl ev mm hm	C	8p11.21	hook homolog 3 (D
FLJ32731	↓	sv pr dl ev mm hm	C	8p11.1	hypothetical protein



Organism: Homo sapiens
Chromosome: 8 **Region Shown:** 39126185.6 40367891.2

Available Maps:
Org: mouse

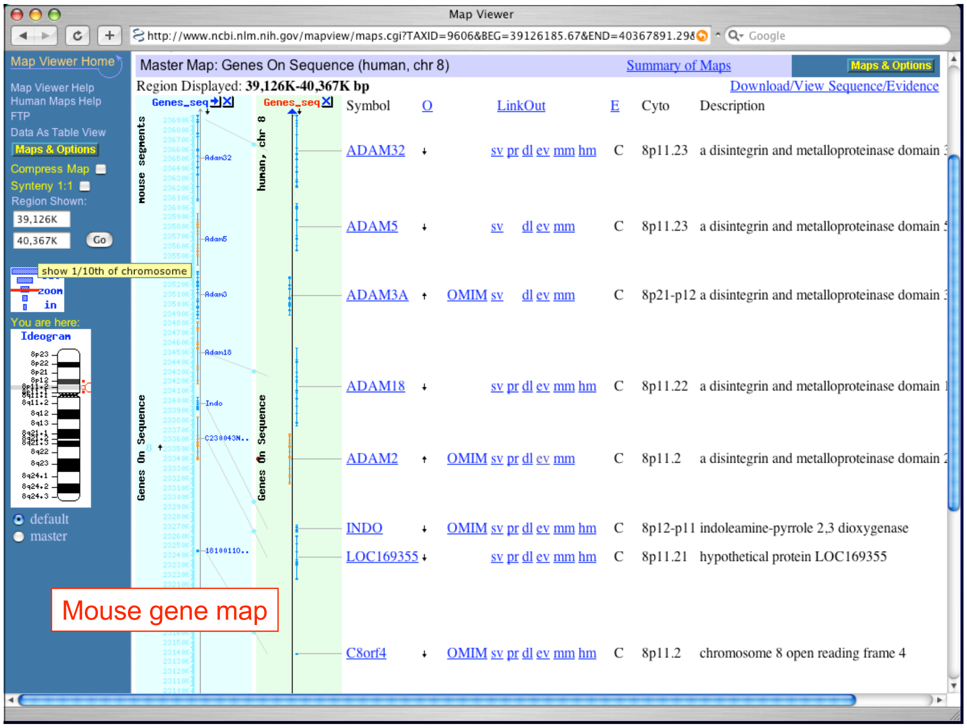
- GenBank_DNA
- Gene
- Hs_EST
- Hs_UniGene
- Mm_EST
- Mm_UniGene
- Rn_EST
- Rn_UniGene
- Phenotype

Maps Displayed (left to right):

- [R][mouse][human:8] Gene
- [ ][human][8] Gene

([R] before map means 'ruler set')

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



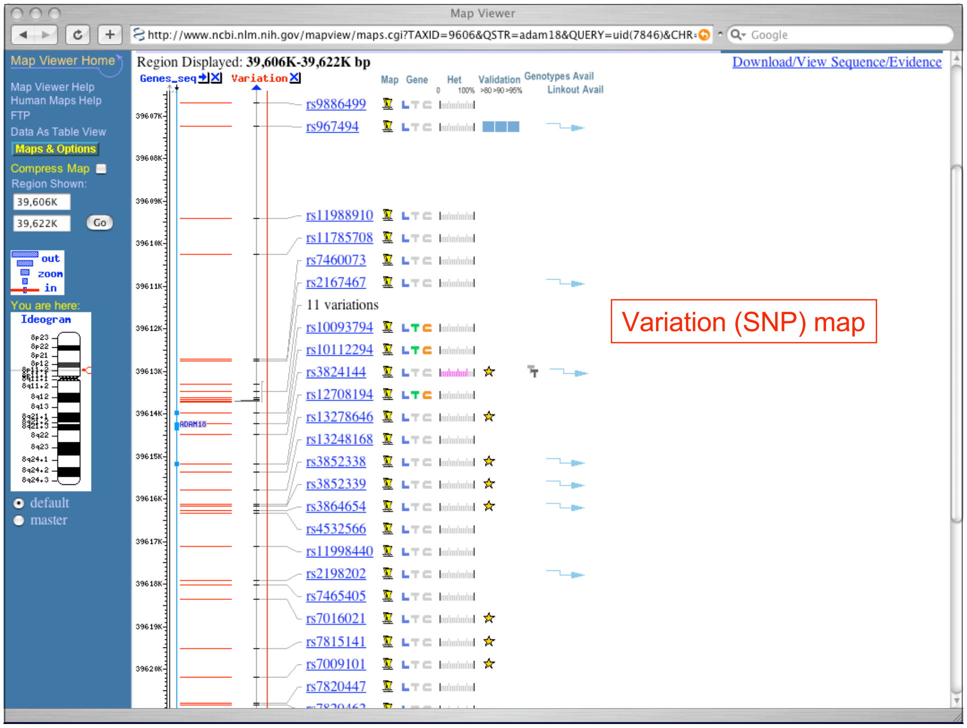
http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?TAXID=9606&BEG=39126185.67&END=40367891.29&

Master Map: Genes On Sequence (human, chr 8)

Region Displayed: **39,126K-40,367K bp**

Symbol	O	LinkOut	E	Cyto	Description
ADAM32 ↓		sv pr dl ev mm hm	C	8p11.23	a disintegrin and metalloproteinase domain
ADAM5 ↓		sv dl ev mm	C	8p11.23	a disintegrin and metalloproteinase domain
ADAM3A ↑	OMIM	sv dl ev mm	C	8p21-p12	a disintegrin and metalloproteinase domain
ADAM18 ↓		sv pr dl ev mm hm	C	8p11.22	a disintegrin and metalloproteinase domain
ADAM2 ↑	OMIM	sv pr dl ev mm	C	8p11.2	a disintegrin and metalloproteinase domain
INDO ↓	OMIM	sv pr dl ev mm hm	C	8p12-p11	indoleamine-pyrrole 2,3 dioxygenase
LOC169355 ↓		sv pr dl ev mm hm	C	8p11.21	hypothetical protein LOC169355
C8orf4 ↓	OMIM	sv pr dl ev mm hm	C	8p11.2	chromosome 8 open reading frame 4

Mouse gene map

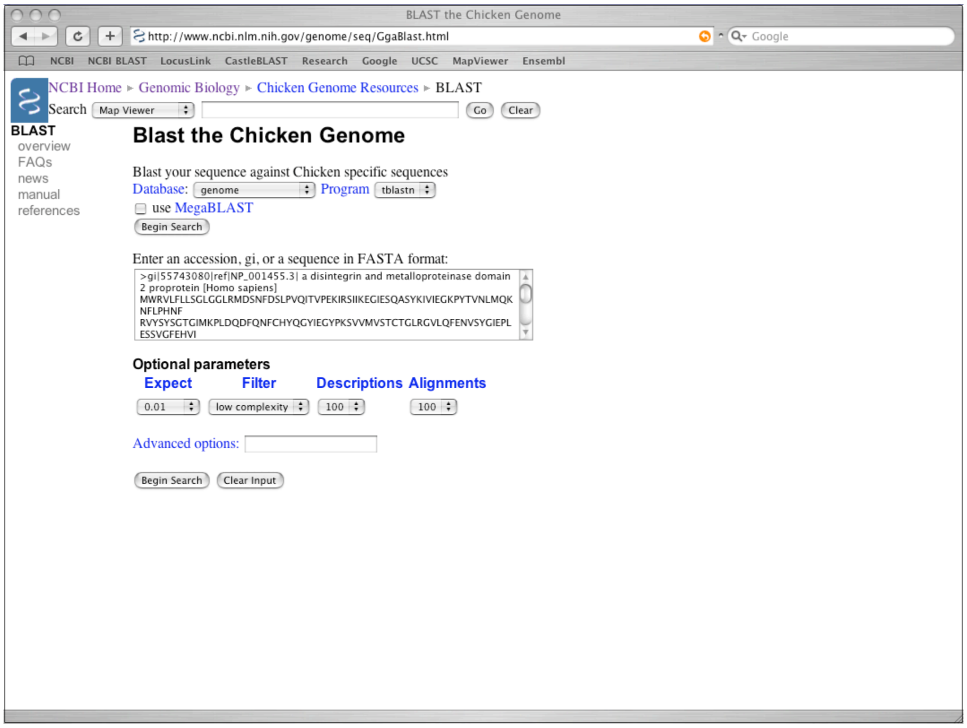


Region Displayed: **39,606K-39,622K bp**

rs9886499
rs967494
rs11988910
rs11785708
rs7460073
rs2167467
11 variations
rs10093794
rs10112294
rs3824144
rs12708194
rs13278646
rs13248168
rs3852338
rs3852339
rs3864654
rs4532566
rs11998440
rs2198202
rs7465405
rs7016021
rs7815141
rs7009101
rs7820447

NCBI

Find a chicken homolog of a human protein



Blast the Chicken Genome

Blast your sequence against Chicken specific sequences

Database: genome
Program tblastn
☐ use MegaBLAST

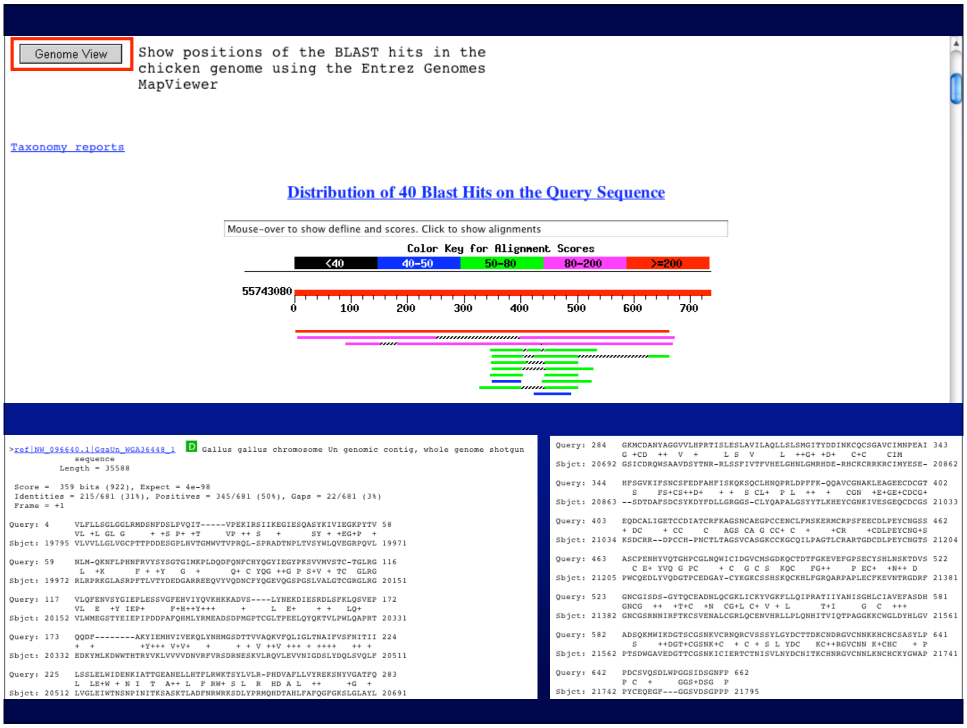
Enter an accession, gi, or a sequence in FASTA format:

```
>gi|55743080|ref|NP_001455.3| a disintegrin and metalloproteinase domain
2 proprotein [Homo sapiens]
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKIVIEGKPYTVNLMQK
NFLPHNF
RVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPL
ESSVGFEHVI
```

Optional parameters

Expect	Filter	Descriptions	Alignments
0.01	low complexity	100	100

Advanced options:

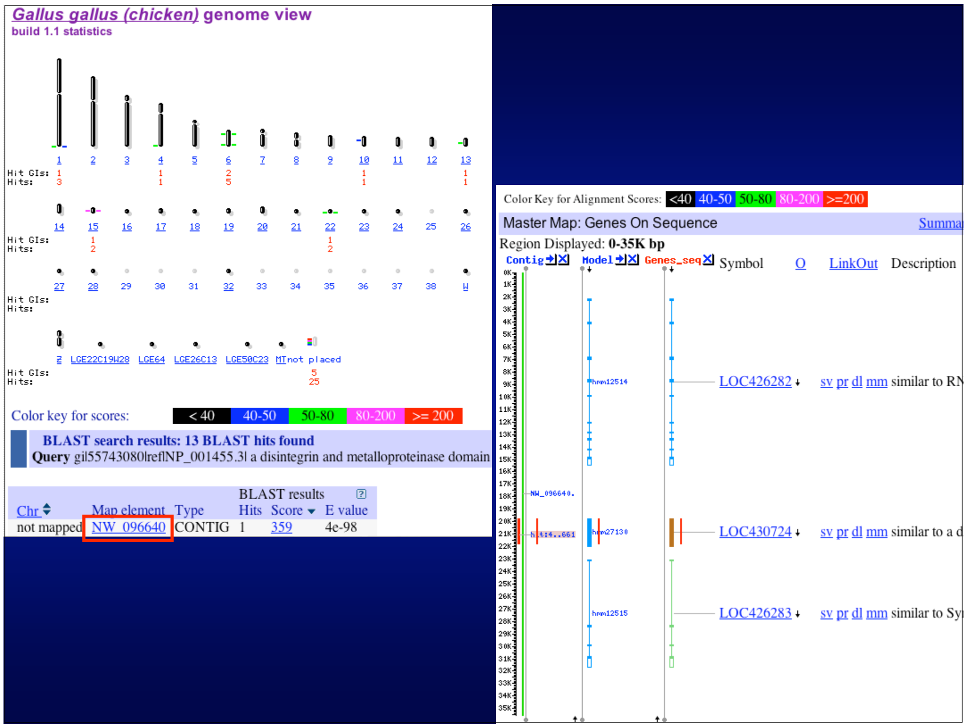


Show positions of the BLAST hits in the chicken genome using the Entrez Genomes MapViewer

Taxonomy reports

Distribution of 40 Blast Hits on the Query Sequence

```
>ref|NW_096640.1|GgaUn_WGA36448_1  D Gallus gallus chromosome Un genomic contig, whole genome shotgun
          sequence
        Length = 35588

 Score =  359 bits (922), Expect = 4e-98
 Identities = 215/681 (31%), Positives = 345/681 (50%), Gaps = 22/681 (3%)
 Frame = +1

Query: 4     VLFLLSGLGGLRMDSNFDSLPVQIT-----VPEKIRSIIKEGIESQASYKIVIEGKPYTV 58
             VL +L GL G     + +S P+ +T     VP ++ S   +       SY + +EG+P  +
Sbjct: 19795 VLVVLLGLVGCPTTPDDESGPLHVTGMWVTVPRQL-SPRADTNPLTVSYWLQVEGRPQVL 19971

Query: 59    NLM-QKNFLPHNFRVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTC-TGLRG 116
              L  +K       F + +Y   G  +       Q+ C YQG ++G P S+V + TC  GLRG
Sbjct: 19972 RLRPRKGLASRPFTLVTYDEDGARREEQVYVQDNCFYQGEVQGSPGSLVALGTCGRGLRG 20151

Query: 117   VLQFENVSYGIEPLESSVGFEHVIYQVKHKKADVS----LYNEKDIESRDLSFKLQSVEP 172
             VL  E  +Y IEP+     F+H++Y+++      +      L  E+     + +   LQ+
Sbjct: 20152 VLWMEGSTYEIEPIPDDPAFQHMLYRMEADSDPMGPTCGLTPEELQYQKTVLPWLQAPRT 20331

Query: 173   QQDF--------AKYIEMHVIVEKQLYNHMGSDTTVVAQKVFQLIGLTNAIFVSFNITII 224
             +  +          +Y+++ V+V+   +      + + V ++V +++ + ++++   ++ +
Sbjct: 20332 EDKYMLKDWWTHTRYVKLVVVVDNVRFVRSDRNESKVLRQVLEVVNIGDSLYDQLSVQLF 20511

Query: 225   LSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLR-PHDVAFLLVYREKSNYVGATFQ 283
             L  LE+W + N I  T  A++ L  F RW+ S L  R  HD A L  ++     +G  +
Sbjct: 20512 LVGLEIWTNSNPINITKSASKTLADFNRWRKSDLYPRMQHDTAHLFAFQGFGKSLGLAYL 20691

Query: 284   GKMCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAI 343
             G +CD  ++  V  +      L S  V      L  ++G+ +D+    C+C     CIM
Sbjct: 20692 GSICDRQWSAAVDSYTNR-RLSSFIVTFVHELGHNLGMRHDE-RHCKCRRKRCIMYESE- 20862

Query: 344   HFSGVKIFSNCSFEDFAHFISKQKSQCLHNQPRLDPFFK-QQAVCGNAKLEAGEECDCGT 402
               S     FS+CS++D+   + +  S CL+  P L  ++  +   CGN  +E+GE+CDCG+
Sbjct: 20863 --SDTDAFSDCSYKDYFDLLGRGGS-CLYQAPALGSYYTLKHEYCGNKIVESGEQCDCGS 21033

Query: 403   EQDCALIGETCCDIATCRFKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSS 462
             + DC    + CC     C   AGS CA G CC+ C  +     +CR     +CDLPEYCNG+S
Sbjct: 21034 KSDCRR--DPCCH-PNCTLTAGSVCASGKCCKGCQILPAGTLCRARTGDCDLPEYCNGTS 21204

Query: 463   ASCPENHYVQTGHPCGLNQWICIDGVCMSGDKQCTDTFGKEVEFGPSECYSHLNSKTDVS 522
               C E+ YVQ G PC     + C  G C S  KQC   FG++     P EC+  +N++ D
Sbjct: 21205 PWCQEDLYVQDGTPCEDGAY-CYKGKCSSHSKQCKHLFGRQARPAPLECFKEVNTRGDRF 21381

Query: 523   GNCGISDS-GYTQCEADNLQCGKLICKYVGKFLLQIPRATIIYANISGHLCIAVEFASDH 581
             GNCG  ++  +T+C  +N  CG+L C+ V + L      T+I      G  C  +++
Sbjct: 21382 GNCGSRNNIRFTKCSVENALCGRLQCENVHRLLPLQNHITVIQTPAGGKKCWGLDYHLGV 21561

Query: 582   ADSQKMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTDKCNDRGVCNNKKHCHCSASYLP 641
               S     ++DGT+CGSNK+C   + C + S L YDC   KC++RGVCNN K+CHC    + P
Sbjct: 21562 PTSDWGAVEDGTTCGSNKICIERTCTNISVLNYDCNITKCHNRGVCNNLKNCHCKYGWAP 21741

Query: 642   PDCSVQSDLWPGGSIDSGNFP 662
             P C  +     GGS+DSG  P
Sbjct: 21742 PYCEQEGF---GGSVDSGPPP 21795
```



Gallus gallus (chicken) genome view

build 1.1 statistics

Color key for scores: < 40 | 40-50 | 50-80 | 80-200 | >= 200

BLAST search results: 13 BLAST hits found

Query gi|55743080|ref|NP_001455.3| a disintegrin and metalloproteinase domain

			BLAST results		
Chr	Map element	Type	Hits	Score	E value
not mapped	NW_096640	CONTIG	1	359	4e-98

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

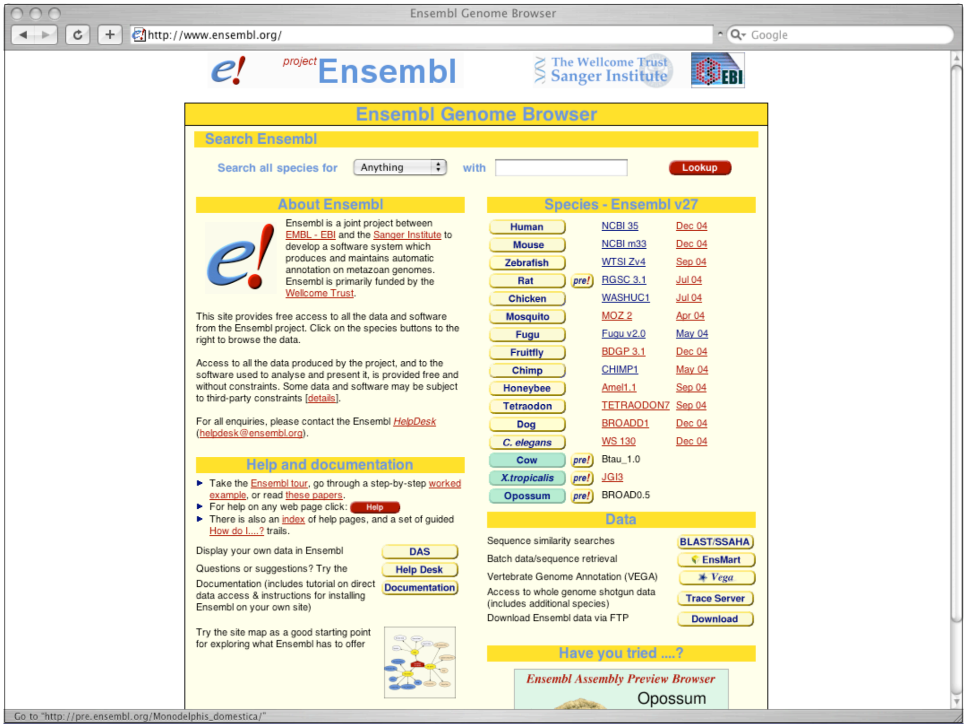
Master Map: Genes On Sequence

Region Displayed: **0-35K bp**

Symbol	O	LinkOut	Description
LOC426282		sv pr dl mm	similar to RN
LOC430724		sv pr dl mm	similar to a d
LOC426283		sv pr dl mm	similar to Sy

Ensembl

Identify genes and SNPs in a chromosomal band



Ensembl Genome Browser

Search Ensembl

About Ensembl

Ensembl is a joint project between EMBL - EBI and the Sanger Institute to develop a software system which produces and maintains automatic annotation on metazoan genomes. Ensembl is primarily funded by the Wellcome Trust.

This site provides free access to all the data and software from the Ensembl project. Click on the species buttons to the right to browse the data.

Access to all the data produced by the project, and to the software used to analyse and present it, is provided free and without constraints. Some data and software may be subject to third-party constraints [details].

For all enquiries, please contact the Ensembl *HelpDesk* (helpdesk@ensembl.org).

Help and documentation

- Take the Ensembl tour, go through a step-by-step worked example, or read these papers.
- For help on any web page click: Help
- There is also an index of help pages, and a set of guided How do I....? trails.

Display your own data in Ensembl

Questions or suggestions? Try the

Documentation (includes tutorial on direct data access & instructions for installing Ensembl on your own site)

Try the site map as a good starting point for exploring what Ensembl has to offer

Species - Ensembl v27

Human		NCBI 35	Dec 04
Mouse		NCBI m33	Dec 04
Zebrafish		WTSI Zv4	Sep 04
Rat	pre!	RGSC 3.1	Jul 04
Chicken		WASHUC1	Jul 04
Mosquito		MOZ 2	Apr 04
Fugu		Fugu v2.0	May 04
Fruitfly		BDGP 3.1	Dec 04
Chimp		CHIMP1	May 04
Honeybee		Amel1.1	Sep 04
Tetraodon		TETRAODON7	Sep 04
Dog		BROADD1	Dec 04
C. elegans		WS 130	Dec 04
Cow	pre!	Btau_1.0	
X.tropicalis	pre!	JGI3	
Opossum	pre!	BROAD0.5	

Data

Sequence similarity searches

Batch data/sequence retrieval

Vertebrate Genome Annotation (VEGA)

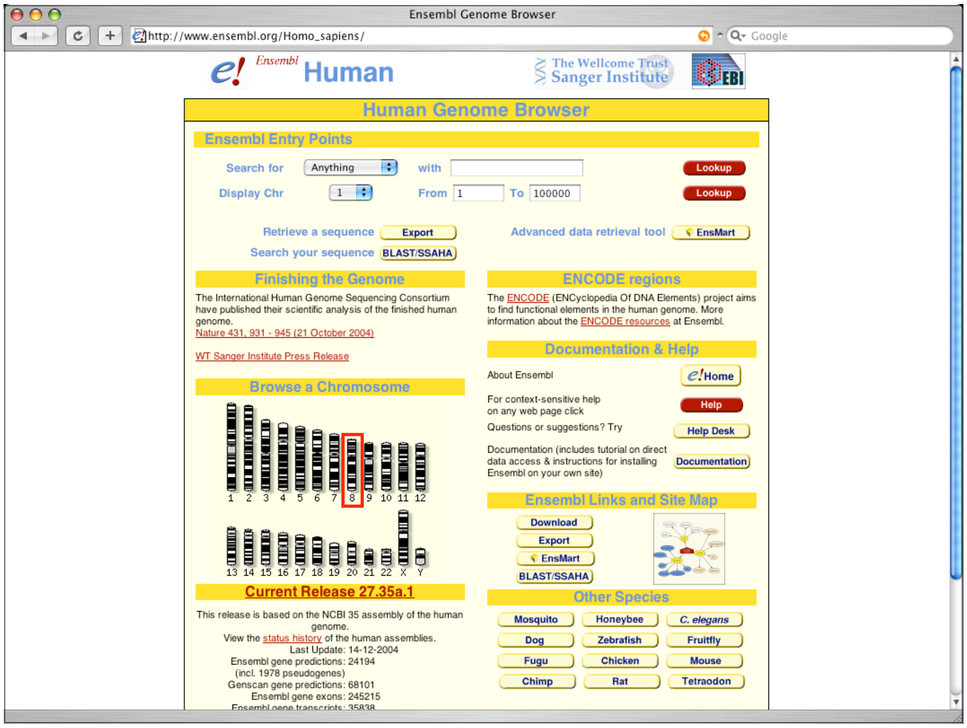
Access to whole genome shotgun data (includes additional species)

Download Ensembl data via FTP

Have you tried ....?

Ensembl Assembly Preview Browser

Opossum



Ensembl Human

Human Genome Browser

Ensembl Entry Points

Retrieve a sequence

Search your sequence

Advanced data retrieval tool

Finishing the Genome

The International Human Genome Sequencing Consortium have published their scientific analysis of the finished human genome.
Nature 431, 931 - 945 (21 October 2004)

WT Sanger Institute Press Release

Browse a Chromosome

Current Release 27.35a.1

This release is based on the NCBI 35 assembly of the human genome.
View the status history of the human assemblies.
Last Update: 14-12-2004
Ensembl gene predictions: 24194
(incl. 1978 pseudogenes)
Genscan gene predictions: 68101
Ensembl gene exons: 245215
Ensembl gene transcripts: 35838

ENCODE regions

The ENCODE (ENCyclopedia Of DNA Elements) project aims to find functional elements in the human genome. More information about the ENCODE resources at Ensembl.

Documentation & Help

About Ensembl

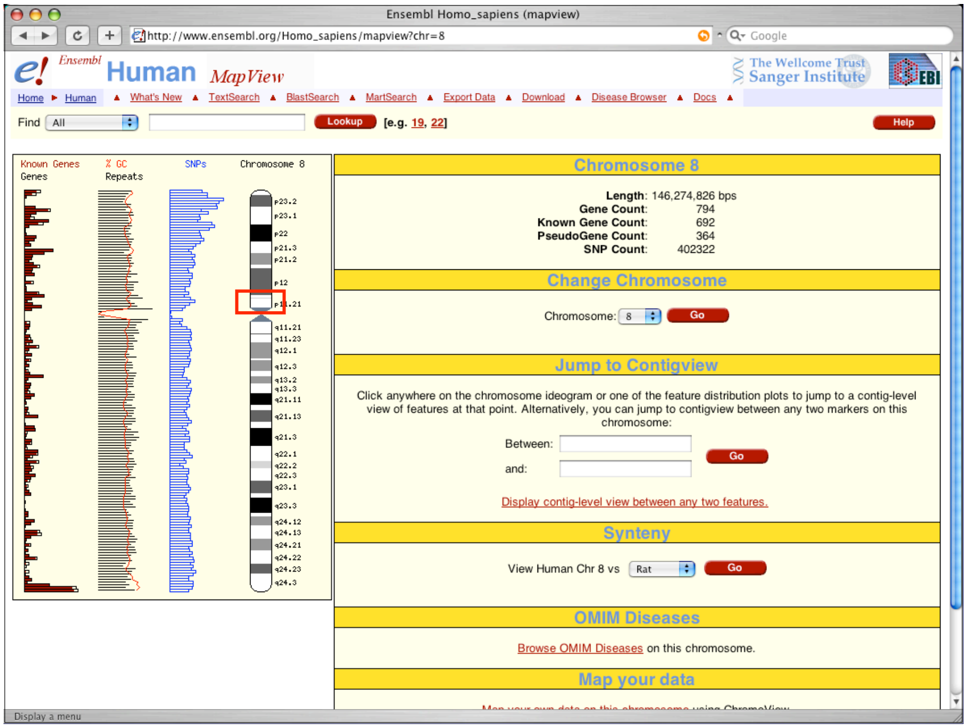
For context-sensitive help on any web page click

Questions or suggestions? Try

Documentation (includes tutorial on direct data access & instructions for installing Ensembl on your own site)

Ensembl Links and Site Map

Other Species



Chromosome 8

Length: 146,274,826 bps
Gene Count: 794
Known Gene Count: 692
PseudoGene Count: 364
SNP Count: 402322

Change Chromosome

Chromosome: 8

Jump to Contigview

Click anywhere on the chromosome ideogram or one of the feature distribution plots to jump to a contig-level view of features at that point. Alternatively, you can jump to contigview between any two markers on this chromosome:

Between:
and:

Display contig-level view between any two features.

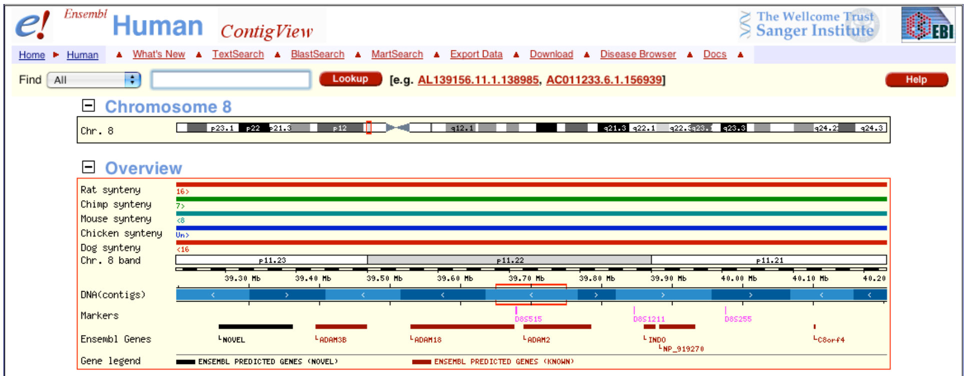
Synteny

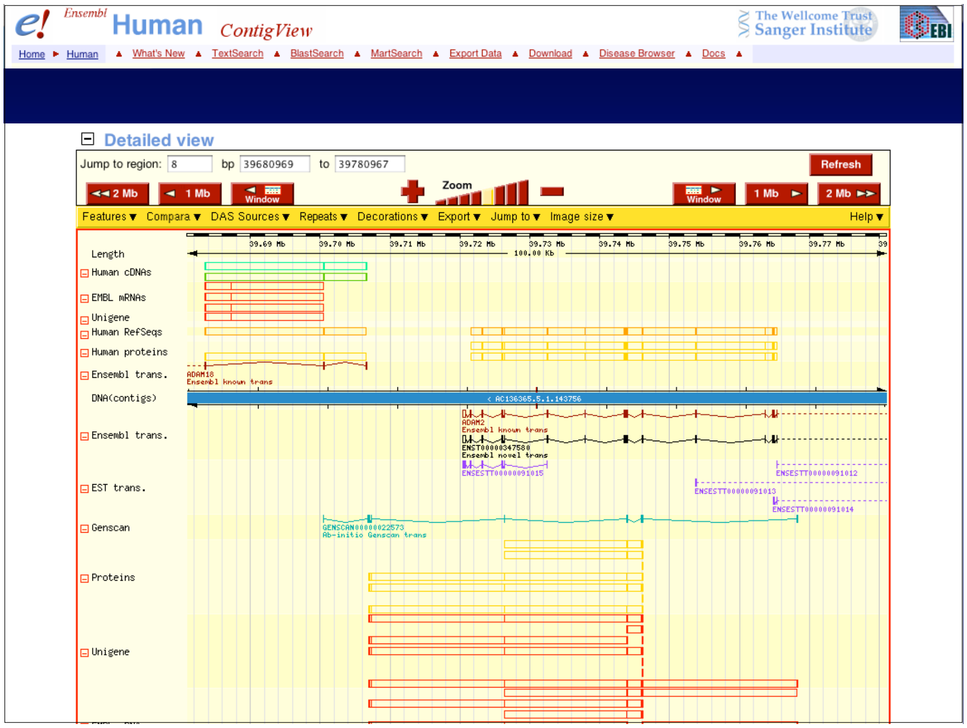
View Human Chr 8 vs Rat

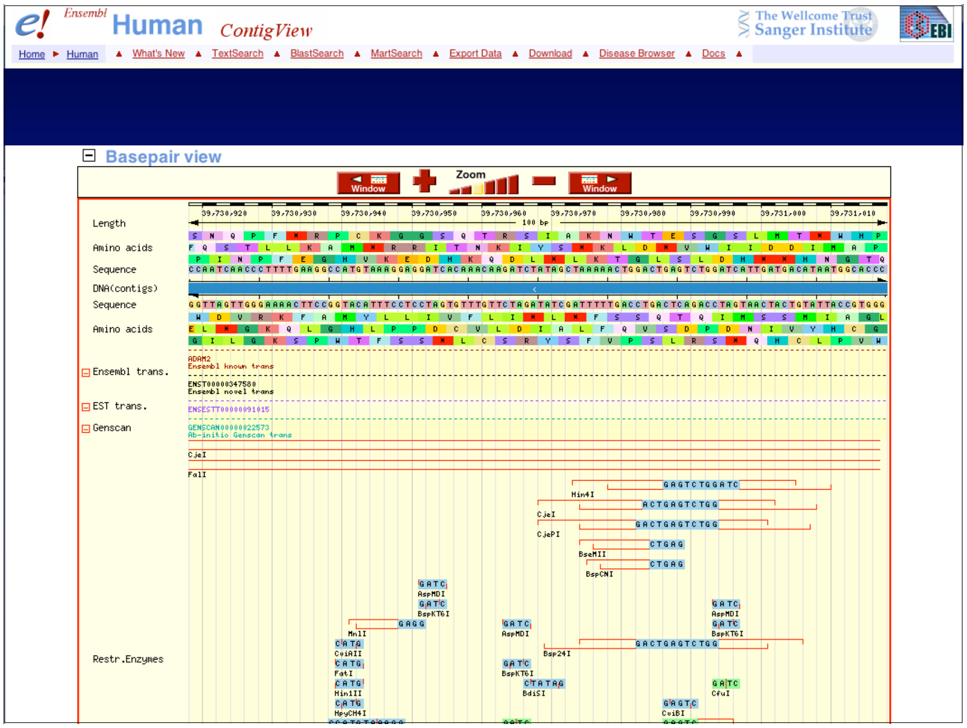
OMIM Diseases

Browse OMIM Diseases on this chromosome.

Map your data

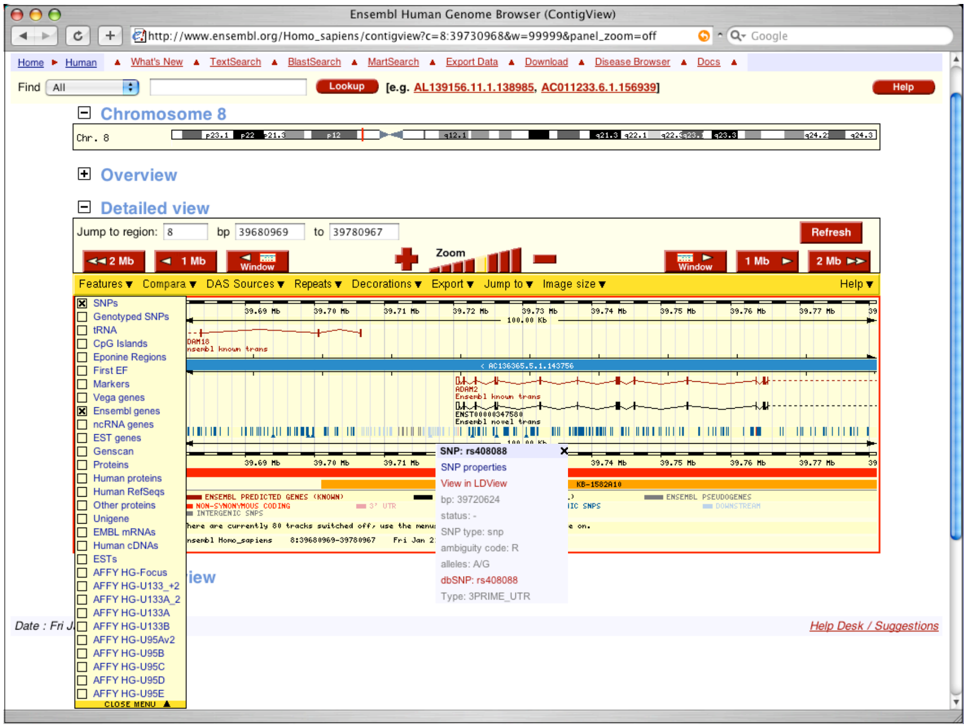



Ensembl
Human
ContigView
The Wellcome Trust Sanger Institute
EBI
Home ▸ Human ▴ What's New ▴ TextSearch ▴ BlastSearch ▴ MartSearch ▴ Export Data ▴ Download ▴ Disease Browser ▴ Docs ▴
Detailed view
Jump to region: 8 bp 39680969 to 39780967
Refresh
◄◄ 2 Mb
◄ 1 Mb
◄ Window
Zoom
Window ►
1 Mb ►
2 Mb ►►
Features ▾ Compara ▾ DAS Sources ▾ Repeats ▾ Decorations ▾ Export ▾ Jump to ▾ Image size ▾
Help ▾
39.69 Mb
39.70 Mb
39.71 Mb
39.72 Mb
39.73 Mb
39.74 Mb
39.75 Mb
39.76 Mb
39.77 Mb
100.00 Kb
Length
Human cDNAs
EMBL mRNAs
Unigene
Human RefSeqs
Human proteins
Ensembl trans.
ADAM18
Ensembl known trans
DNA(contigs)
< AC136365.5.1.143756
Ensembl trans.
ADAM2
Ensembl known trans
ENST00000347580
Ensembl novel trans
ENSESTT00000091015
ENSESTT00000091012
EST trans.
ENSESTT00000091013
ENSESTT00000091014
Genscan
GENSCAN00000022573
Ab-initio Genscan trans
Proteins
Unigene


Ensembl
Human
ContigView
The Wellcome Trust Sanger Institute
EBI
Home ▸ Human ▴ What's New ▴ TextSearch ▴ BlastSearch ▴ MartSearch ▴ Export Data ▴ Download ▴ Disease Browser ▴ Docs ▴
Basepair view
◄ Window
Zoom
Window ►
39,730,920
39,730,930
39,730,940
39,730,950
39,730,960
39,730,970
39,730,980
39,730,990
39,731,000
39,731,010
100 bp
Length
Amino acids
Sequence
DNA(contigs)
Sequence
Amino acids
Ensembl trans.
ADAM2
Ensembl known trans
ENST00000347580
Ensembl novel trans
EST trans.
ENSESTT00000091015
Genscan
GENSCAN00000022573
Ab-initio Genscan trans
CjeI
FalI
Hin4I
CjeI
CjePI
BseMII
BspCNI
AspMDI
BspKT6I
MnlI
CviAII
FatI
Hin1II
HpyCH4I
AspMDI
BspKT6I
Bsp24I
BdiSI
CfuI
CviBI
Restr.Enzymes

Ensembl

Change the tracks displayed on the ContigView



Ensembl Human Genome Browser (ContigView)

http://www.ensembl.org/Homo_sapiens/contigview?c=8:39730968&w=99999&panel_zoom=off

Find: All

[e.g. AL139156.11.1.138985, AC011233.6.1.156939]

Jump to region: 8 bp 39680969 to 39780967

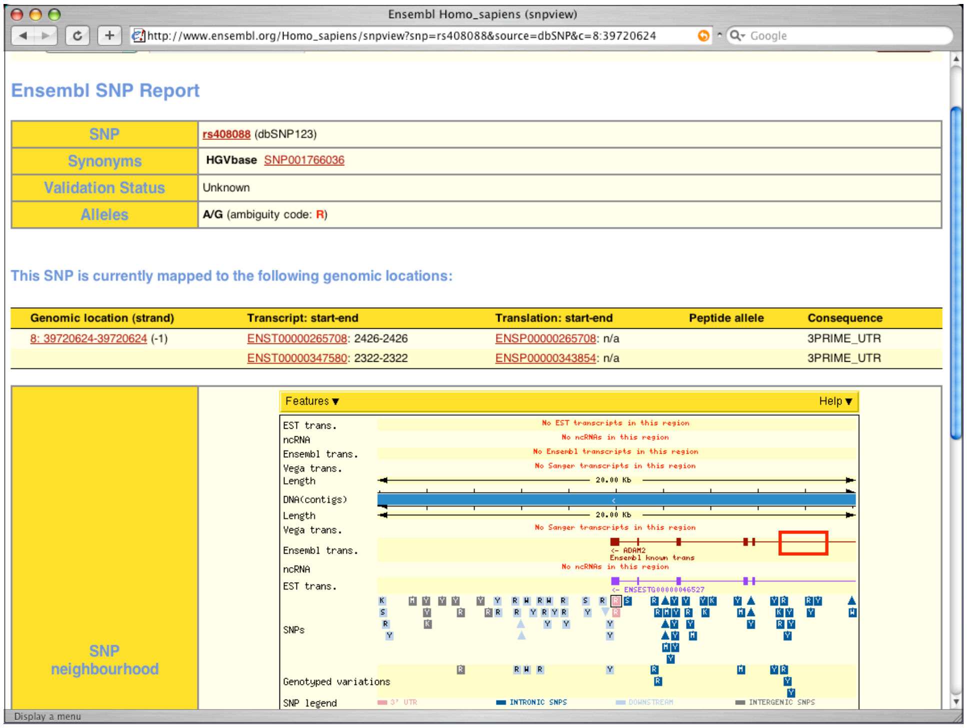
SNP: rs408088
SNP properties
View in LDView
bp: 39720624
status: -
SNP type: snp
ambiguity code: R
alleles: A/G
dbSNP: rs408088
Type: 3PRIME_UTR

here are currently 80 tracks switched off, use the menus … e on.

nsembl Homo_sapiens 8:39680969-39780967 Fri Jan 2…

Date : Fri J…

Help Desk / Suggestions



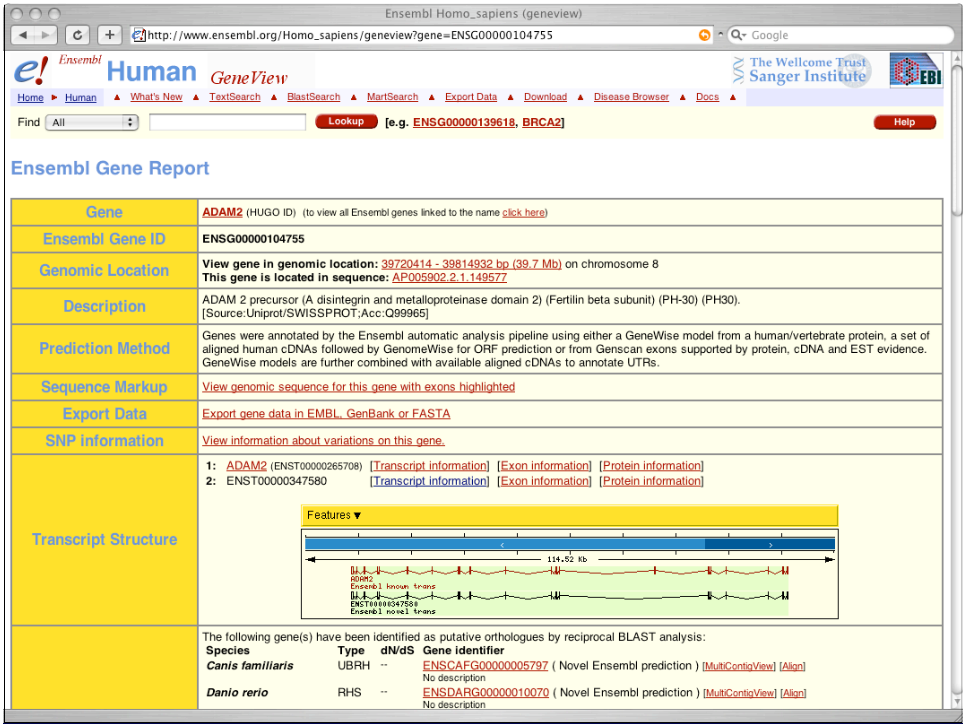
Ensembl SNP Report

SNP	rs408088 (dbSNP123)
Synonyms	**HGVbase** SNP001766036
Validation Status	Unknown
Alleles	**A/G** (ambiguity code: **R**)

This SNP is currently mapped to the following genomic locations:

Genomic location (strand)	Transcript: start-end	Translation: start-end	Peptide allele	Consequence
8: 39720624-39720624 (-1)	ENST00000265708: 2426-2426	ENSP00000265708: n/a		3PRIME_UTR
	ENST00000347580: 2322-2322	ENSP00000343854: n/a		3PRIME_UTR

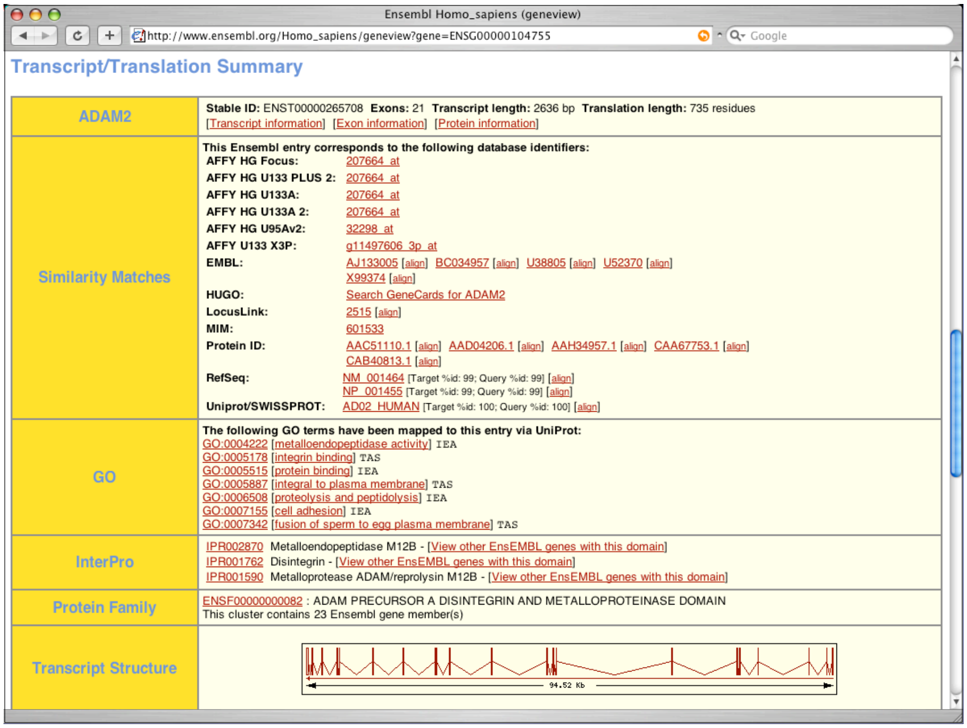
SNP neighbourhood



Ensembl Gene Report

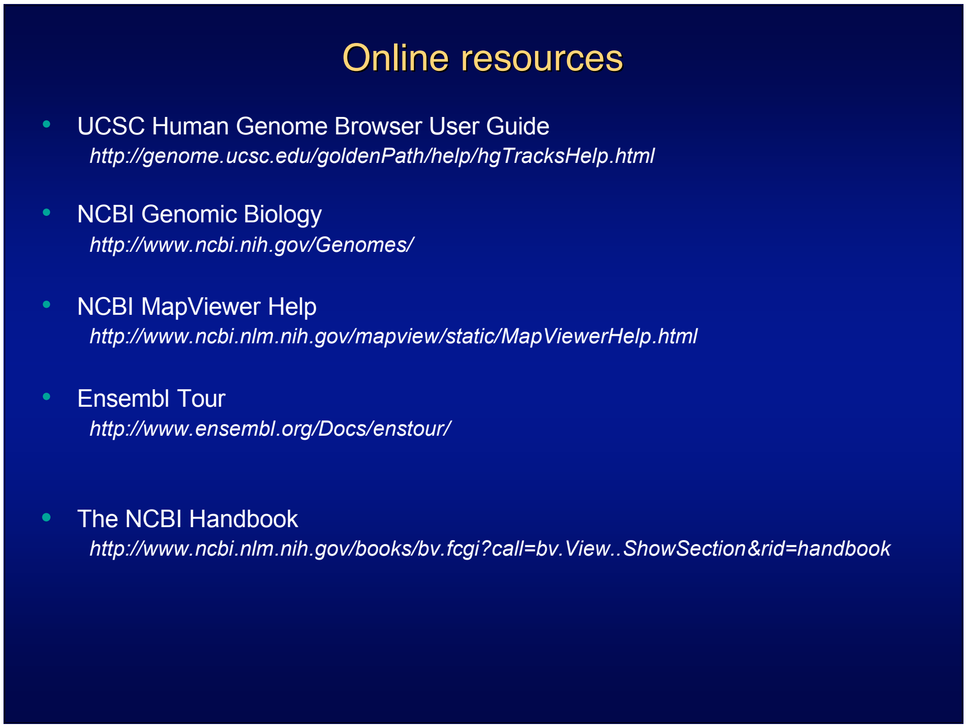
Gene	ADAM2 (HUGO ID) (to view all Ensembl genes linked to the name click here)
Ensembl Gene ID	**ENSG00000104755**
Genomic Location	**View gene in genomic location:** 39720414 - 39814932 bp (39.7 Mb) on chromosome 8 **This gene is located in sequence:** AP005902.2.1.149577
Description	ADAM 2 precursor (A disintegrin and metalloproteinase domain 2) (Fertilin beta subunit) (PH-30) (PH30). [Source:Uniprot/SWISSPROT;Acc:Q99965]
Prediction Method	Genes were annotated by the Ensembl automatic analysis pipeline using either a GeneWise model from a human/vertebrate protein, a set of aligned human cDNAs followed by GenomeWise for ORF prediction or from Genscan exons supported by protein, cDNA and EST evidence. GeneWise models are further combined with available aligned cDNAs to annotate UTRs.
Sequence Markup	View genomic sequence for this gene with exons highlighted
Export Data	Export gene data in EMBL, GenBank or FASTA
SNP information	View information about variations on this gene.
Transcript Structure	**1:** ADAM2 (ENST00000265708) [Transcript information] [Exon information] [Protein information] **2:** ENST00000347580 [Transcript information] [Exon information] [Protein information]
	The following gene(s) have been identified as putative orthologues by reciprocal BLAST analysis:

Species	Type	dN/dS	Gene identifier
Canis familiaris	UBRH	--	ENSCAFG00000005797 (Novel Ensembl prediction) [MultiContigView] [Align] No description
Danio rerio	RHS	--	ENSDARG00000010070 (Novel Ensembl prediction) [MultiContigView] [Align] No description



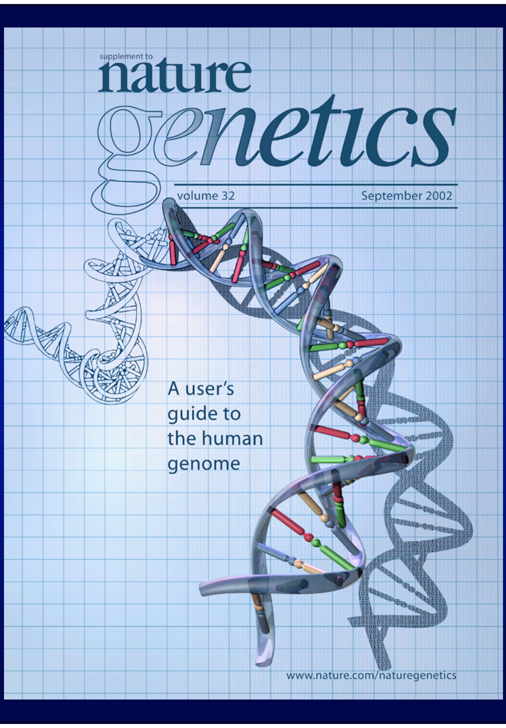
Transcript/Translation Summary

ADAM2	**Stable ID:** ENST00000265708 **Exons:** 21 **Transcript length:** 2636 bp **Translation length:** 735 residues [Transcript information] [Exon information] [Protein information]
Similarity Matches	**This Ensembl entry corresponds to the following database identifiers:** **AFFY HG Focus:** 207664_at **AFFY HG U133 PLUS 2:** 207664_at **AFFY HG U133A:** 207664_at **AFFY HG U133A 2:** 207664_at **AFFY HG U95Av2:** 32298_at **AFFY U133 X3P:** g11497606_3p_at **EMBL:** AJ133005 [align] BC034957 [align] U38805 [align] U52370 [align] X99374 [align] **HUGO:** Search GeneCards for ADAM2 **LocusLink:** 2515 [align] **MIM:** 601533 **Protein ID:** AAC51110.1 [align] AAD04206.1 [align] AAH34957.1 [align] CAA67753.1 [align] CAB40813.1 [align] **RefSeq:** NM_001464 [Target %id: 99; Query %id: 99] [align] NP_001455 [Target %id: 99; Query %id: 99] [align] **Uniprot/SWISSPROT:** AD02_HUMAN [Target %id: 100; Query %id: 100] [align]
GO	**The following GO terms have been mapped to this entry via UniProt:** GO:0004222 [metalloendopeptidase activity] IEA GO:0005178 [integrin binding] TAS GO:0005515 [protein binding] IEA GO:0005887 [integral to plasma membrane] TAS GO:0006508 [proteolysis and peptidolysis] IEA GO:0007155 [cell adhesion] IEA GO:0007342 [fusion of sperm to egg plasma membrane] TAS
InterPro	IPR002870 Metalloendopeptidase M12B - [View other EnsEMBL genes with this domain] IPR001762 Disintegrin - [View other EnsEMBL genes with this domain] IPR001590 Metalloprotease ADAM/reprolysin M12B - [View other EnsEMBL genes with this domain]
Protein Family	ENSF00000000082 : ADAM PRECURSOR A DISINTEGRIN AND METALLOPROTEINASE DOMAIN This cluster contains 23 Ensembl gene member(s)
Transcript Structure	94.52 Kb



Online 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 Tour
  http://www.ensembl.org/Docs/enstour/
- The NCBI Handbook
  http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowSection&rid=handbook



supplement to
nature genetics
volume 32 September 2002

A user's guide to the human genome

www.nature.com/naturegenetics

http://www.nature.com/genomics/

References

- Current Protocols in Bioinformatics
  - *UNIT 1.4: The UCSC Genome Browser*
  - *UNIT 1.5: Using the NCBI Map Viewer to Browse Genomic Sequence Data*
  - Access through *http://nihlibrary.nih.gov/ResearchTools/OnlineJournals.htm*
- UCSC
  - Hsu F *et al*. The UCSC Proteome Browser. Nucleic Acids Res. 2005. 33:D454-8.
  - Karolchik D *et al*. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004. 32:D493-6.
  - Karolchik D *et al*. The UCSC Genome Browser Database. Nucleic Acids Res. 2003. 31:51-4.
- NCBI
  - Wheeler DL *et al*. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2005:D39-45.
- Ensembl
  - Hubbard T *et al*. Ensembl 2005. Nucleic Acids Res. 2005. 33:D447-53.
  - Hammond MP, and Birney E. Genome information resources - developments at Ensembl. 2004. Trends Genet. 20:268-72.
  - Birney E *et al*. An overview of Ensembl. 2004. Genome Res. 14:925-8.