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

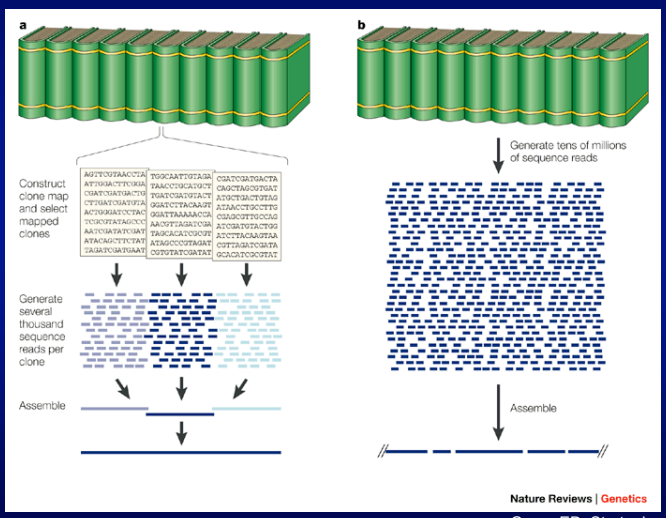
Types of data integrated in genome browsers

- 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



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

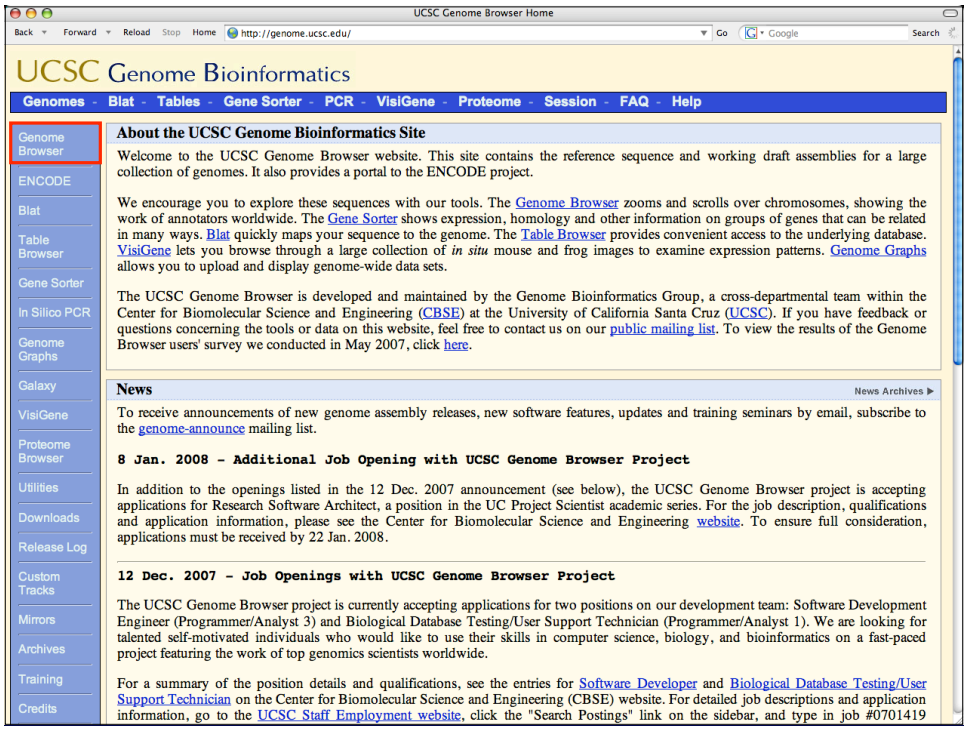
http://www.ncbi.nlm.nih.gov/RefSeq/key.html

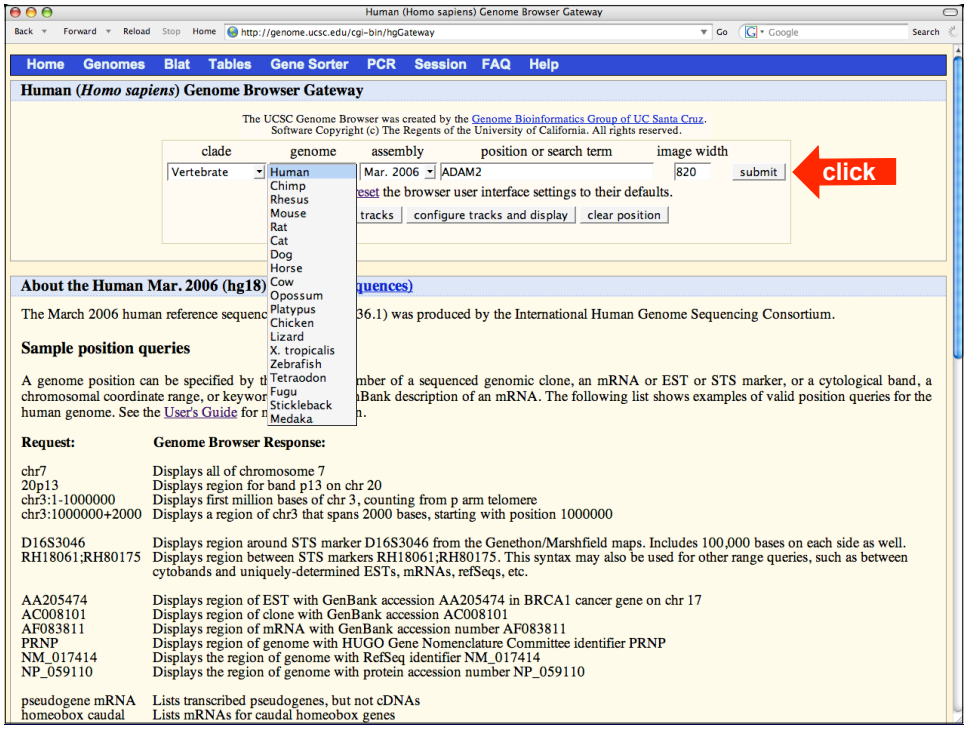
Beta actin mRNA RefSeq

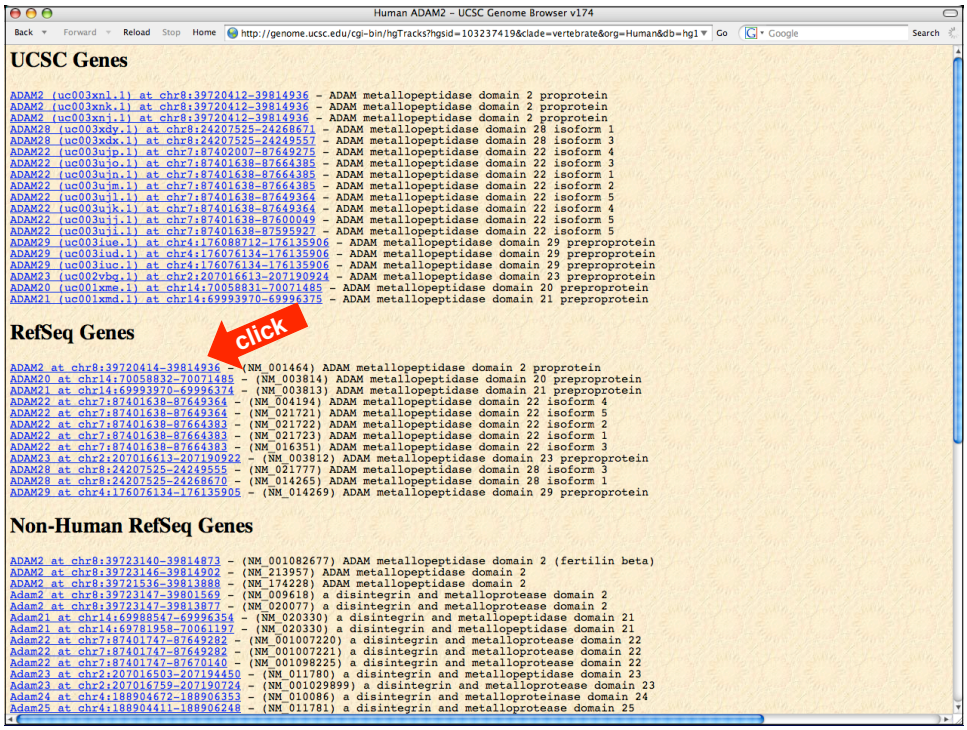
```
LOCUS       NM_001101               1793 bp    mRNA    linear   PRI 14-JAN-2008
DEFINITION  Homo sapiens actin, beta (ACTB), mRNA.
ACCESSION   NM_001101
VERSION     NM_001101.2  GI:5016088
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1793)
  AUTHORS   Villebeck,L., Moparthi,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   17939680
                                     .....

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:4501884.

            Summary: This gene encodes one of six different actin proteins.
            Actins are highly conserved proteins that are involved in cell
            motility, structure, and integrity. This actin is a major
            constituent of the contractile apparatus and one of the two
            nonmuscle cytoskeletal actins.

            Publication Note:  This RefSeq record includes a subset of the
            publications that are available for this gene. Please see the
            Entrez Gene record to access additional publications.
            COMPLETENESS: complete on the 3' end.
                                     .....

CDS             74..1201
                /gene="ACTB"
                /GO_component="actin filament; cytoskeleton; TIP60 histone
                acetyltransferase complex [PMID 10966108]"
                /GO_function="ATP binding; nucleotide binding; protein
                binding [PMID 15527767]; structural constituent of
                cytoskeleton"
                /note="beta cytoskeletal actin; PS1TP5-binding protein 1;
                actin, cytoplasmic 1; beta-actin"
                /codon_start=1
                /product="beta actin"
                /protein_id="NP_001092.1"
                /db_xref="GI:4501885"
                /db_xref="CCDS:CCDS5341.1"
                /db_xref="GeneID:60"
                /db_xref="HGNC:132"
                /db_xref="HPRD:00032"
                /db_xref="MIM:102630"
                /translation="MDDDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVM
                VGMGQKDSYVGDEAQSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHP
                VLLTEAPLNPKANREKMTQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVT
                HTVPIYEGYALPHAILRLDLAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCY
                VALDFEQEMATAASSSSLEKSYELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHE
                TTFNSIMKCDVDIRKDLYANTVLSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPE
                RKYSVWIGGSILASLSTFQQMWISKQEYDESGPSIVHRKCF"
                                     .....

ORIGIN
        1 cgcgtccgcc ccgcgagcac agagcctcgc ctttgccgat ccgccgcccg tccacacccg
       61 ccgccagctc accatggatg atgatatcgc cgcgctcgtc gtcgacaacg gctccggcat
      121 gtgcaaggcc ggcttcgcgg gcgacgatgc ccccgggcc gtcttcccct ccatcgtggg
```

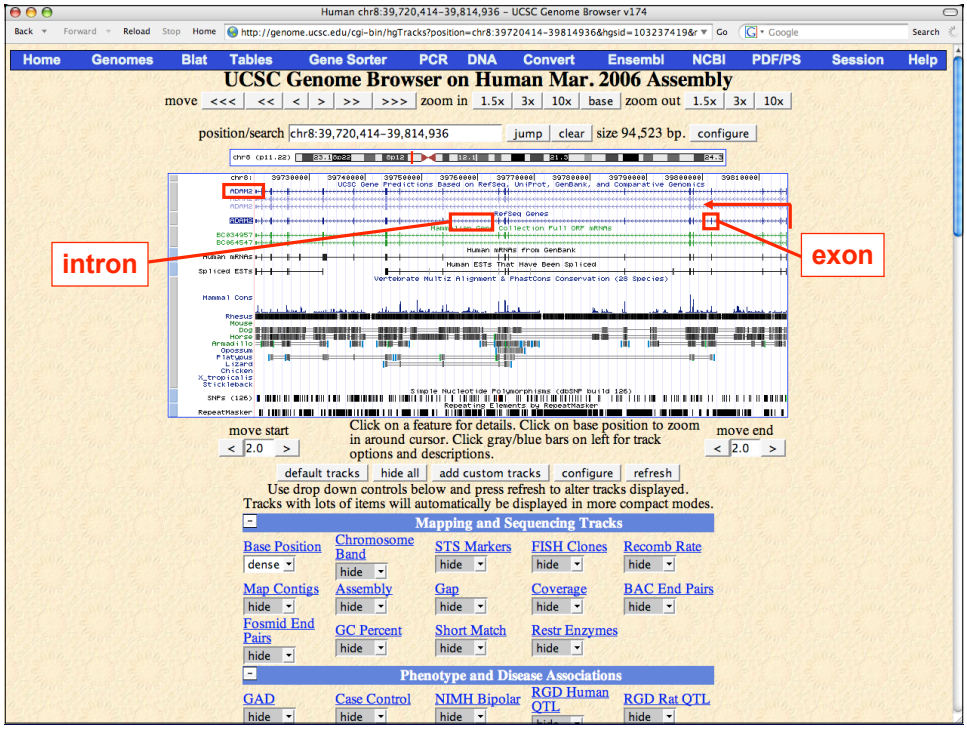
UCSC

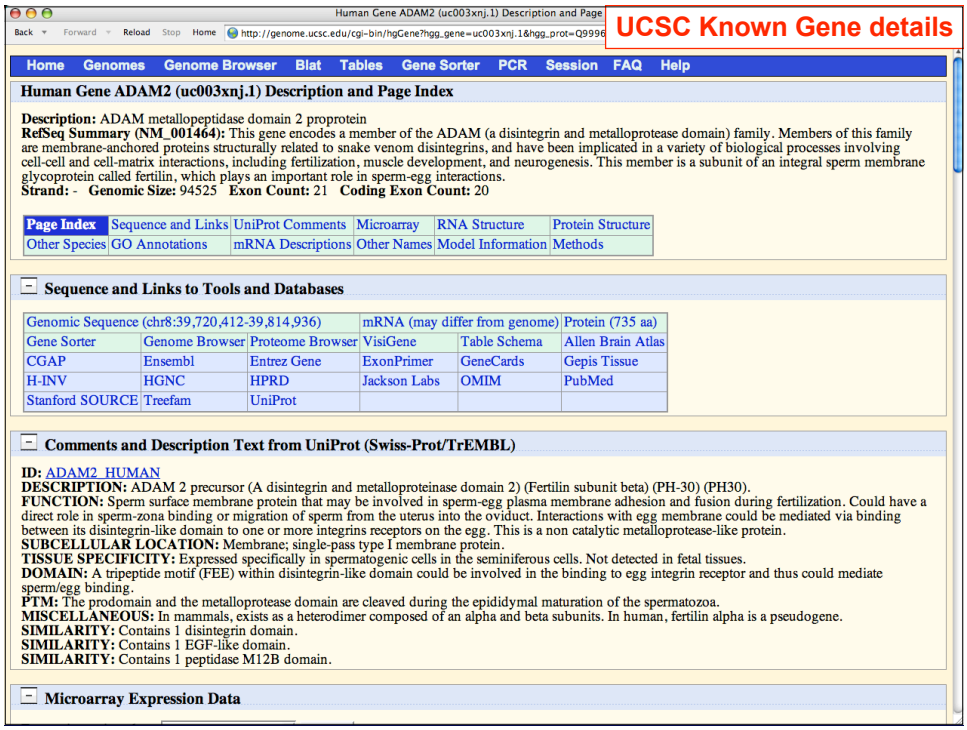
View a region in the genome by querying with a gene symbol

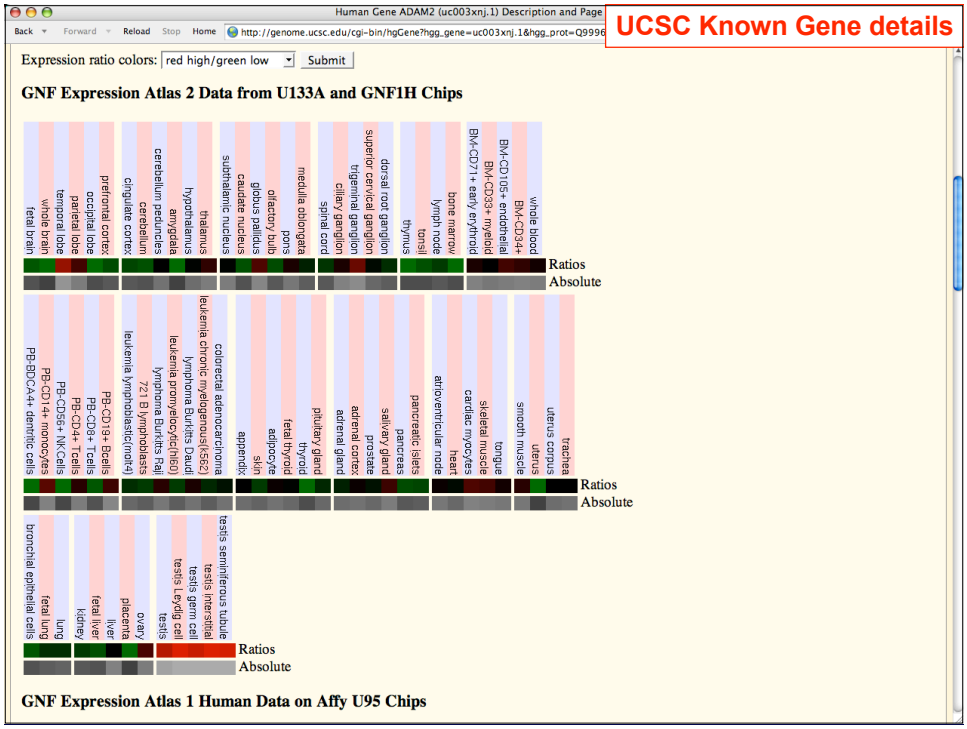


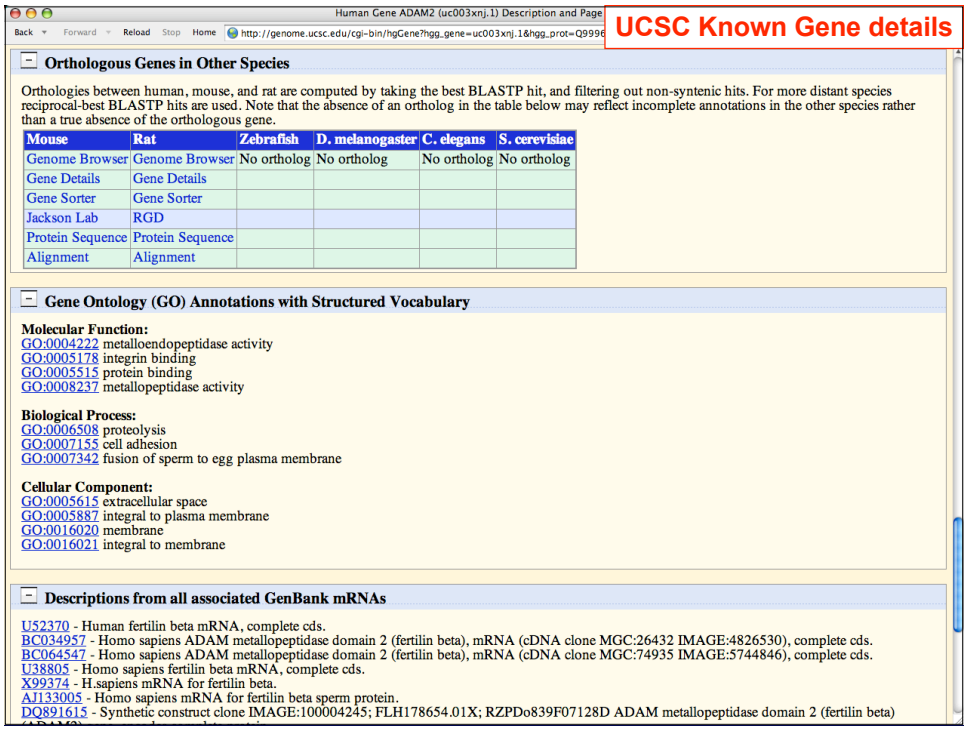

Human (Homo sapiens) Genome Browser Gateway
http://genome.ucsc.edu/cgi-bin/hgGateway
Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help
Human (Homo sapiens) Genome Browser Gateway
The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.
clade
genome
assembly
position or search term
image width
Vertebrate
Human
Chimp
Rhesus
Mouse
Rat
Cat
Dog
Horse
Cow
Opossum
Platypus
Chicken
Lizard
X. tropicalis
Zebrafish
Tetraodon
Fugu
Stickleback
Medaka
Mar. 2006
ADAM2
820
submit
click
reset the browser user interface settings to their defaults.
tracks
configure tracks and display
clear position
About the Human Mar. 2006 (hg18)
Sample position queries
Request:
Genome Browser Response:
chr7
Displays all of chromosome 7
20p13
Displays region for band p13 on chr 20
chr3:1-1000000
Displays first million bases of chr 3, counting from p arm telomere
chr3:1000000+2000
Displays a region of chr3 that spans 2000 bases, starting with position 1000000
D16S3046
Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
RH18061;RH80175
Displays region between STS markers RH18061;RH80175. This syntax may also be used for other range queries, such as between cytobands and uniquely-determined ESTs, mRNAs, refSeqs, etc.
AA205474
Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101
Displays region of clone with GenBank accession AC008101
AF083811
Displays region of mRNA with GenBank accession number AF083811
PRNP
Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414
Displays the region of genome with RefSeq identifier NM_017414
NP_059110
Displays the region of genome with protein accession number NP_059110
pseudogene mRNA
Lists transcribed pseudogenes, but not cDNAs
homeobox caudal
Lists mRNAs for caudal homeobox genes

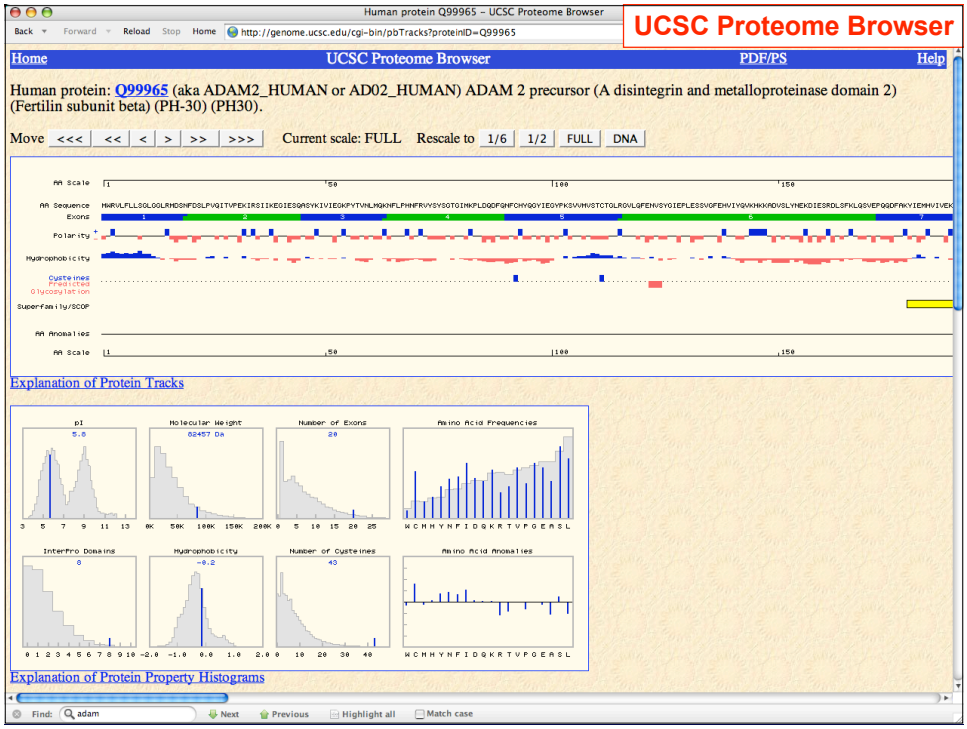

Human ADAM2 - UCSC Genome Browser v174
http://genome.ucsc.edu/cgi-bin/hgTracks?hgsid=103237419&clade=vertebrate&org=Human&db=hg1
UCSC Genes
ADAM2 (uc003xnl.1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein
ADAM2 (uc003xnk.1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein
ADAM2 (uc003xnj.1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein
ADAM28 (uc003xdy.1) at chr8:24207525-24268671 - ADAM metallopeptidase domain 28 isoform 1
ADAM28 (uc003xdx.1) at chr8:24207525-24249557 - ADAM metallopeptidase domain 28 isoform 3
ADAM22 (uc003ujp.1) at chr7:87402007-87649275 - ADAM metallopeptidase domain 22 isoform 4
ADAM22 (uc003ujo.1) at chr7:87401638-87664385 - ADAM metallopeptidase domain 22 isoform 3
ADAM22 (uc003ujn.1) at chr7:87401638-87664385 - ADAM metallopeptidase domain 22 isoform 1
ADAM22 (uc003ujm.1) at chr7:87401638-87664385 - ADAM metallopeptidase domain 22 isoform 2
ADAM22 (uc003ujl.1) at chr7:87401638-87649364 - ADAM metallopeptidase domain 22 isoform 5
ADAM22 (uc003ujk.1) at chr7:87401638-87649364 - ADAM metallopeptidase domain 22 isoform 4
ADAM22 (uc003uji.1) at chr7:87401638-87600049 - ADAM metallopeptidase domain 22 isoform 5
ADAM22 (uc003ujj.1) at chr7:87401638-87595927 - ADAM metallopeptidase domain 22 isoform 5
ADAM29 (uc003iue.1) at chr4:176088712-176135906 - ADAM metallopeptidase domain 29 preproprotein
ADAM29 (uc003iud.1) at chr4:176076134-176135906 - ADAM metallopeptidase domain 29 preproprotein
ADAM29 (uc003iuc.1) at chr4:176076134-176135906 - ADAM metallopeptidase domain 29 preproprotein
ADAM23 (uc002ybg.1) at chr2:207016613-207190924 - ADAM metallopeptidase domain 23 preproprotein
ADAM20 (uc001xme.1) at chr14:70058831-70071485 - ADAM metallopeptidase domain 20 preproprotein
ADAM21 (uc001xmd.1) at chr14:69993970-69996375 - ADAM metallopeptidase domain 21 preproprotein
RefSeq Genes
click
ADAM2 at chr8:39720414-39814936 - (NM_001464) ADAM metallopeptidase domain 2 proprotein
ADAM20 at chr14:70058832-70071485 - (NM_003814) ADAM metallopeptidase domain 20 preproprotein
ADAM21 at chr14:69993970-69996374 - (NM_003813) ADAM metallopeptidase domain 21 preproprotein
ADAM22 at chr7:87401638-87649364 - (NM_004194) ADAM metallopeptidase domain 22 isoform 4
ADAM22 at chr7:87401638-87649364 - (NM_021721) ADAM metallopeptidase domain 22 isoform 5
ADAM22 at chr7:87401638-87664383 - (NM_021722) ADAM metallopeptidase domain 22 isoform 2
ADAM22 at chr7:87401638-87664383 - (NM_021723) ADAM metallopeptidase domain 22 isoform 1
ADAM22 at chr7:87401638-87664383 - (NM_016351) ADAM metallopeptidase domain 22 isoform 3
ADAM23 at chr2:207016613-207190922 - (NM_003812) ADAM metallopeptidase domain 23 preproprotein
ADAM28 at chr8:24207525-24249555 - (NM_021777) ADAM metallopeptidase domain 28 isoform 3
ADAM28 at chr8:24207525-24268670 - (NM_014265) ADAM metallopeptidase domain 28 isoform 1
ADAM29 at chr4:176076134-176135905 - (NM_014269) ADAM metallopeptidase domain 29 preproprotein
Non-Human RefSeq Genes
ADAM2 at chr8:39723140-39814873 - (NM_001082677) ADAM metallopeptidase domain 2 (fertilin beta)
ADAM2 at chr8:39723146-39814902 - (NM_213957) ADAM metallopeptidase domain 2
ADAM2 at chr8:39721536-39813888 - (NM_174228) ADAM metallopeptidase domain 2
Adam2 at chr8:39723147-39801569 - (NM_009618) a disintegrin and metalloprotease domain 2
Adam2 at chr8:39723147-39813877 - (NM_020077) a disintegrin and metalloprotease domain 2
Adam21 at chr14:69988547-69996354 - (NM_020330) a disintegrin and metallopeptidase domain 21
Adam21 at chr14:69781958-70061197 - (NM_020330) a disintegrin and metallopeptidase domain 21
Adam22 at chr7:87401747-87649282 - (NM_001007220) a disintegrin and metalloprotease domain 22
Adam22 at chr7:87401747-87649282 - (NM_001007221) a disintegrin and metalloprotease domain 22
Adam22 at chr7:87401747-87670140 - (NM_001098225) a disintegrin and metalloprotease domain 22
Adam23 at chr2:207016503-207194450 - (NM_011780) a disintegrin and metallopeptidase domain 23
Adam23 at chr2:207016759-207190724 - (NM_001029899) a disintegrin and metalloprotease domain 23
Adam24 at chr4:188904672-188906353 - (NM_010086) a disintegrin and metalloproteinase domain 24
Adam25 at chr4:188904411-188906248 - (NM_011781) a disintegrin and metalloprotease domain 25

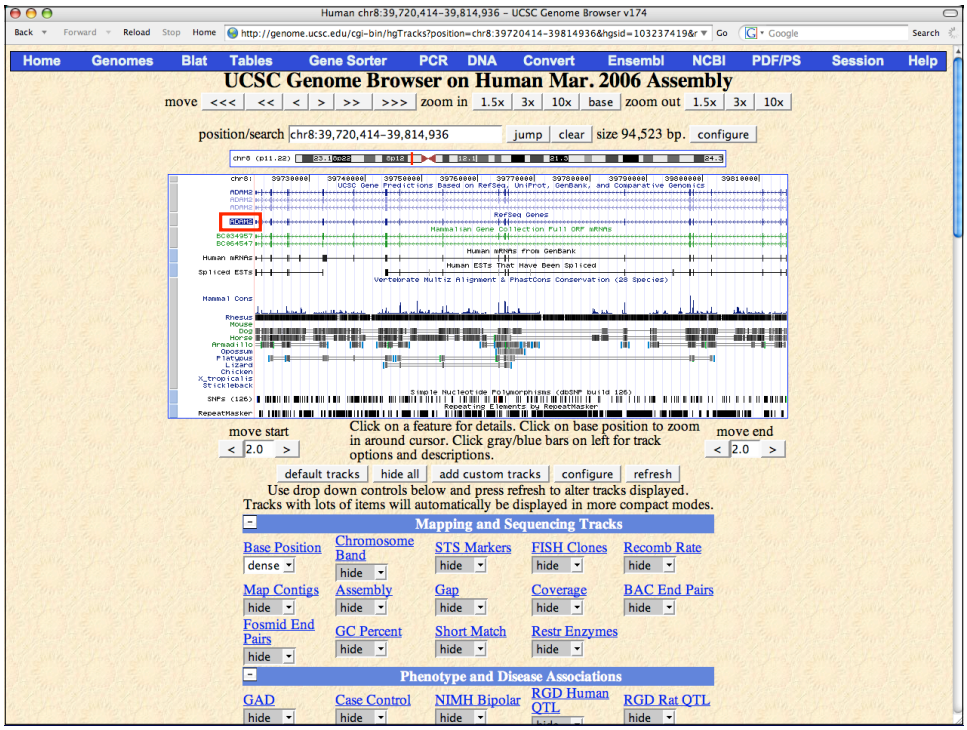

Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v174
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help
UCSC Genome Browser on Human Mar. 2006 Assembly
intron
exon
Mapping and Sequencing Tracks
Phenotype and Disease Associations

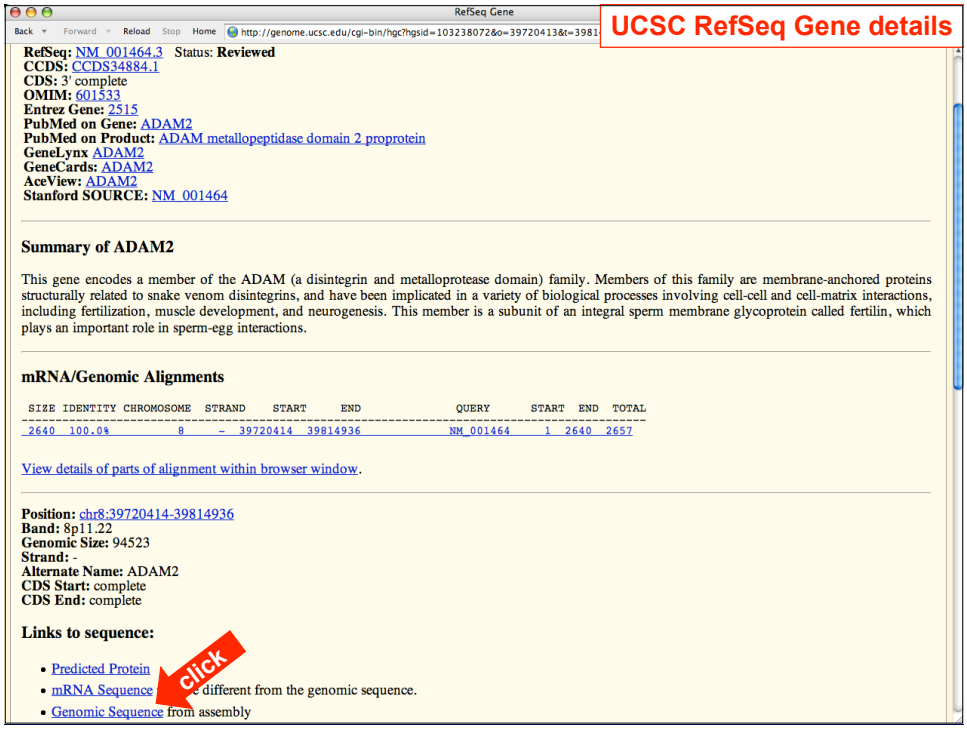

UCSC Known Gene details
Human Gene ADAM2 (uc003xnj.1) Description and Page Index
Sequence and Links to Tools and Databases
Comments and Description Text from UniProt (Swiss-Prot/TrEMBL)
Microarray Expression Data

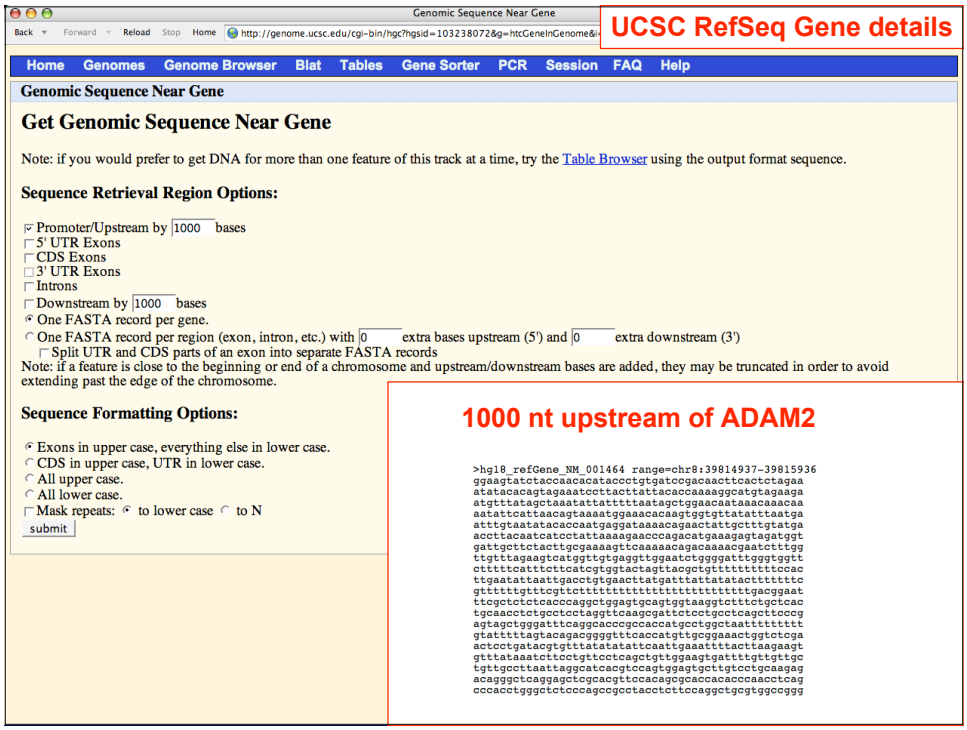

Human Gene ADAM2 (uc003xnj.1) Description and Page
Back Forward Reload Stop Home http://genome.ucsc.edu/cgi-bin/hgGene?hgg_gene=uc003xnj.1&hgg_prot=Q9996
UCSC Known Gene details
Expression ratio colors: red high/green low Submit
GNF Expression Atlas 2 Data from U133A and GNF1H Chips
fetal brain
whole brain
temporal lobe
parietal lobe
occipital lobe
prefrontal cortex
cingulate cortex
cerebellum
cerebellum peduncles
amygdala
hypothalamus
thalamus
subthalamic nucleus
caudate nucleus
globus pallidus
olfactory bulb
pons
medulla oblongata
spinal cord
ciliary ganglion
trigeminal ganglion
superior cervical ganglion
dorsal root ganglion
thymus
tonsil
lymph node
bone marrow
BM-CD71+ early erythroid
BM-CD33+ myeloid
BM-CD105+ endothelial
BM-CD34+
whole blood
Ratios
Absolute
PB-BDCA4+ dendritic cells
PB-CD14+ monocytes
PB-CD56+ NK Cells
PB-CD4+ Tcells
PB-CD8+ Tcells
PB-CD19+ Bcells
leukemia lymphoblastic(mol+)
721.B lymphoblasts
lymphoma Burkitts Raji
leukemia promyelocytic(HL60)
lymphoma Burkitts Daudi
leukemia chronic myelogenous(K562)
colorectal adenocarcinoma
appendix
skin
adipocyte
fetal thyroid
thyroid
pituitary gland
adrenal gland
adrenal cortex
prostate
salivary gland
pancreas
pancreatic islets
atrioventricular node
heart
cardiac myocytes
skeletal muscle
tongue
smooth muscle
uterus
uterus corpus
trachea
Ratios
Absolute
bronchial epithelial cells
fetal lung
lung
kidney
fetal liver
liver
placenta
ovary
testis
testis Leydig cell
testis germ cell
testis interstitial
testis seminiferous tubule
Ratios
Absolute
GNF Expression Atlas 1 Human Data on Affy U95 Chips


Human Gene ADAM2 (uc003xnj.1) Description and Page
Back Forward Reload Stop Home http://genome.ucsc.edu/cgi-bin/hgGene?hgg_gene=uc003xnj.1&hgg_prot=Q9996
UCSC Known Gene details
Orthologous Genes in Other Species
Orthologies between human, mouse, and rat are computed by taking the best BLASTP hit, and filtering out non-syntenic hits. For more distant species reciprocal-best BLASTP hits are used. Note that the absence of an ortholog in the table below may reflect incomplete annotations in the other species rather than a true absence of the orthologous gene.
Mouse | Rat | Zebrafish | D. melanogaster | C. elegans | S. cerevisiae
Genome Browser | Genome Browser | No ortholog | No ortholog | No ortholog | No ortholog
Gene Details | Gene Details
Gene Sorter | Gene Sorter
Jackson Lab | RGD
Protein Sequence | Protein Sequence
Alignment | Alignment
Gene Ontology (GO) Annotations with Structured Vocabulary
Molecular Function:
GO:0004222 metalloendopeptidase activity
GO:0005178 integrin binding
GO:0005515 protein binding
GO:0008237 metallopeptidase activity
Biological Process:
GO:0006508 proteolysis
GO:0007155 cell adhesion
GO:0007342 fusion of sperm to egg plasma membrane
Cellular Component:
GO:0005615 extracellular space
GO:0005887 integral to plasma membrane
GO:0016020 membrane
GO:0016021 integral to membrane
Descriptions from all associated GenBank mRNAs
U52370 - Human fertilin beta mRNA, complete cds.
BC034957 - Homo sapiens ADAM metallopeptidase domain 2 (fertilin beta), mRNA (cDNA clone MGC:26432 IMAGE:4826530), complete cds.
BC064547 - Homo sapiens ADAM metallopeptidase domain 2 (fertilin beta), mRNA (cDNA clone MGC:74935 IMAGE:5744846), complete cds.
U38805 - Homo sapiens fertilin beta mRNA, complete cds.
X99374 - H.sapiens mRNA for fertilin beta.
AJ133005 - Homo sapiens mRNA for fertilin beta sperm protein.
DQ891615 - Synthetic construct clone IMAGE:100004245; FLH178654.01X; RZPDo839F07128D ADAM metallopeptidase domain 2 (fertilin beta)


UCSC Proteome Browser
Human protein Q99965 – UCSC Proteome Browser
http://genome.ucsc.edu/cgi-bin/pbTracks?proteinID=Q99965
Home
UCSC Proteome Browser
PDF/PS
Help
Human protein: Q99965 (aka ADAM2_HUMAN or AD02_HUMAN) ADAM 2 precursor (A disintegrin and metalloproteinase domain 2) (Fertilin subunit beta) (PH-30) (PH30).
Move
Current scale: FULL
Rescale to
Explanation of Protein Tracks
Explanation of Protein Property Histograms

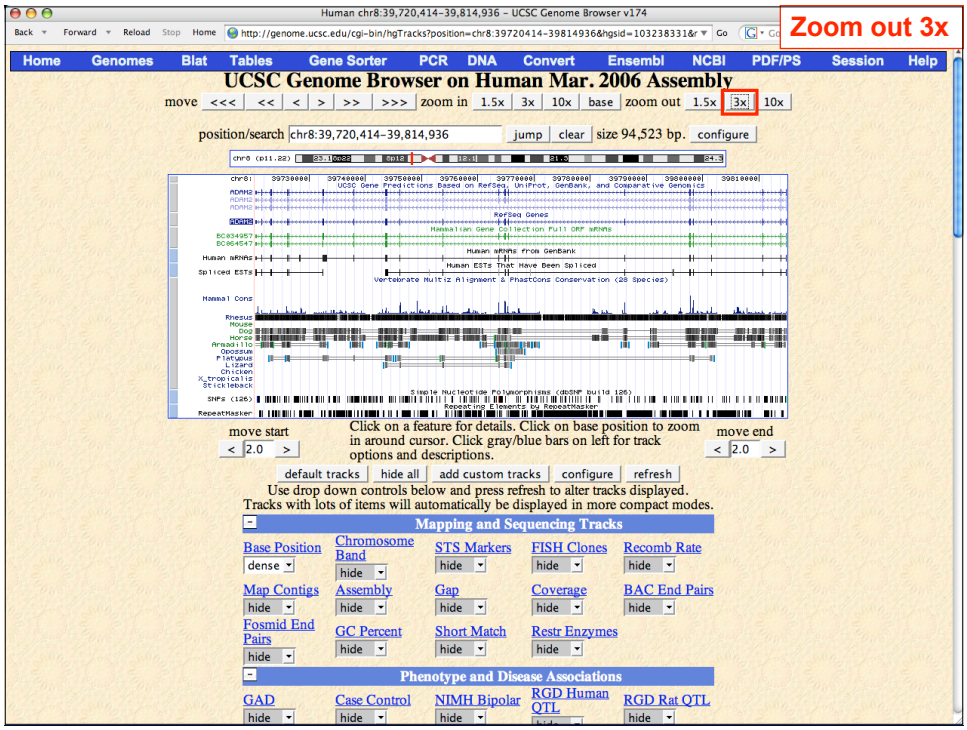

Human chr8:39,720,414-39,814,936 – UCSC Genome Browser v174
Home
Genomes
Blat
Tables
Gene Sorter
PCR
DNA
Convert
Ensembl
NCBI
PDF/PS
Session
Help
UCSC Genome Browser on Human Mar. 2006 Assembly
move
zoom in
zoom out
position/search
chr8:39,720,414-39,814,936
jump
clear
size 94,523 bp.
configure
move start
move end
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
Map Contigs
Assembly
Gap
Coverage
BAC End Pairs
Fosmid End Pairs
GC Percent
Short Match
Restr Enzymes
Phenotype and Disease Associations
GAD
Case Control
NIMH Bipolar
RGD Human QTL
RGD Rat QTL

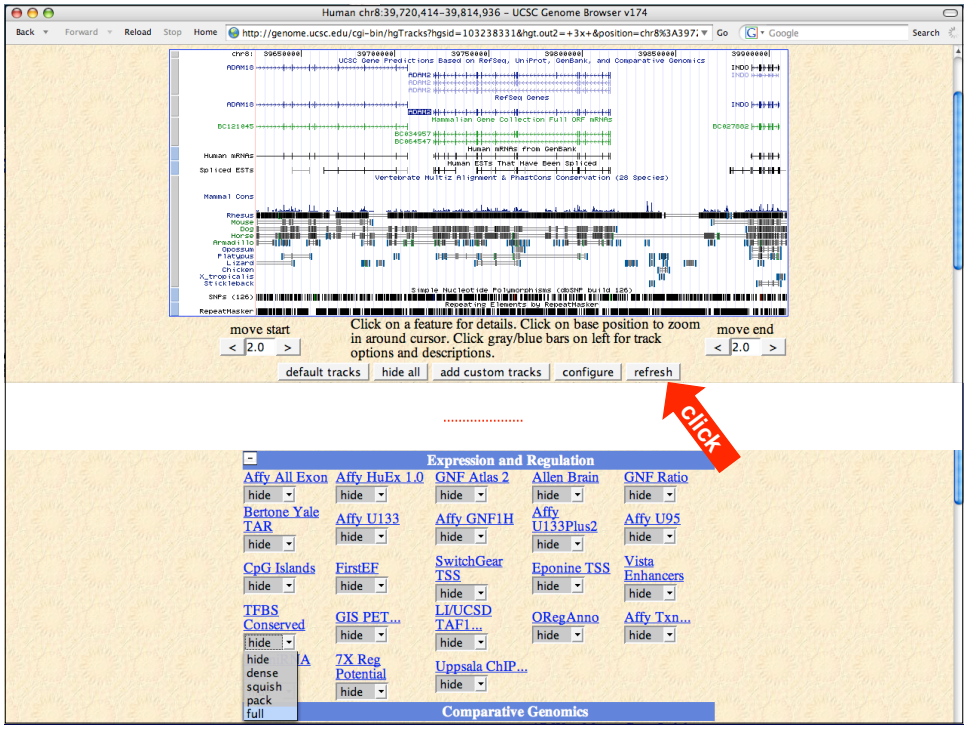

RefSeq Gene
UCSC RefSeq Gene details
RefSeq: NM_001464.3 Status: Reviewed
CCDS: CCDS34884.1
CDS: 3' complete
OMIM: 601533
Entrez Gene: 2515
PubMed on Gene: ADAM2
PubMed on Product: ADAM metallopeptidase domain 2 proprotein
GeneLynx ADAM2
GeneCards: ADAM2
AceView: ADAM2
Stanford SOURCE: NM_001464
Summary of ADAM2
This gene encodes a member of the ADAM (a disintegrin and metalloprotease domain) family. Members of this family are membrane-anchored proteins structurally related to snake venom disintegrins, and have been implicated in a variety of biological processes involving cell-cell and cell-matrix interactions, including fertilization, muscle development, and neurogenesis. This member is a subunit of an integral sperm membrane glycoprotein called fertilin, which plays an important role in sperm-egg interactions.
mRNA/Genomic Alignments
SIZE IDENTITY CHROMOSOME STRAND START END QUERY START END TOTAL
2640 100.0% 8 - 39720414 39814936 NM_001464 1 2640 2657
View details of parts of alignment within browser window.
Position: chr8:39720414-39814936
Band: 8p11.22
Genomic Size: 94523
Strand: -
Alternate Name: ADAM2
CDS Start: complete
CDS End: complete
Links to sequence:
Predicted Protein
mRNA Sequence
different from the genomic sequence.
Genomic Sequence from assembly
Click

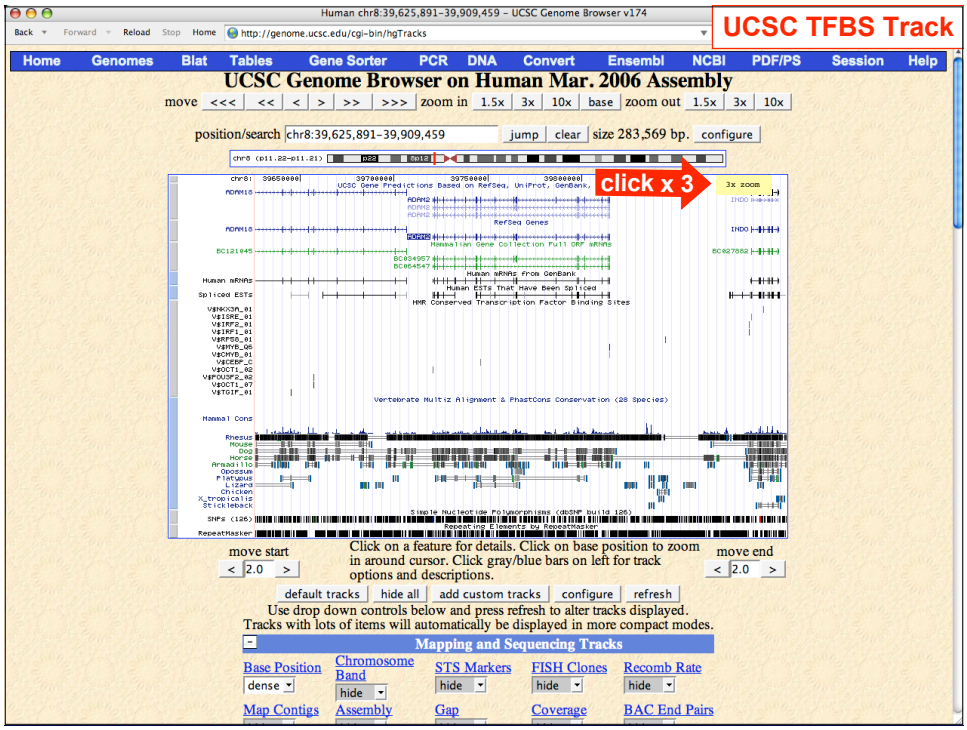

Genomic Sequence Near Gene
UCSC RefSeq Gene details
Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help
Genomic Sequence Near Gene
Get Genomic Sequence Near Gene
Note: if you would prefer to get DNA for more than one feature of this track at a time, try the Table Browser using the output format sequence.
Sequence Retrieval Region Options:
Promoter/Upstream by 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
1000 nt upstream of ADAM2
>hg18_refGene_NM_001464 range=chr8:39814937-39815936

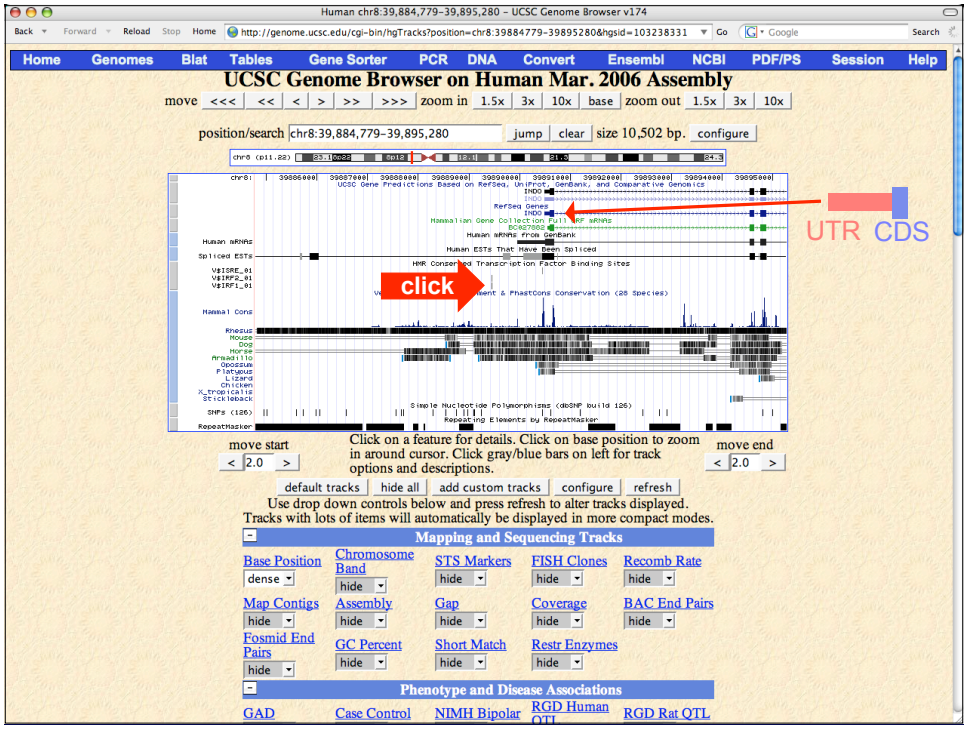
UCSC

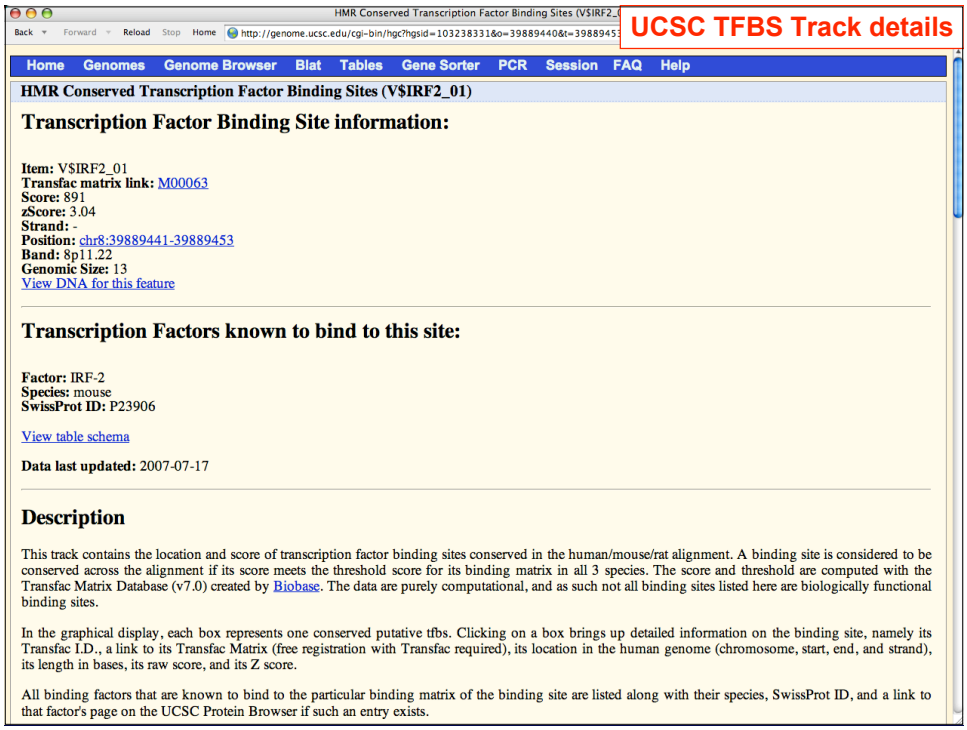
Add tracks to the Genome Browser




Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v174
move start
move end
Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.
default tracks
hide all
add custom tracks
configure
refresh
click
Expression and Regulation
Affy All Exon
Affy HuEx 1.0
GNF Atlas 2
Allen Brain
GNF Ratio
Bertone Yale TAR
Affy U133
Affy GNF1H
Affy U133Plus2
Affy U95
CpG Islands
FirstEF
SwitchGear TSS
Eponine TSS
Vista Enhancers
TFBS Conserved
GIS PET...
LI/UCSD TAF1...
ORegAnno
Affy Txn...
7X Reg Potential
Uppsala ChIP...
Comparative Genomics

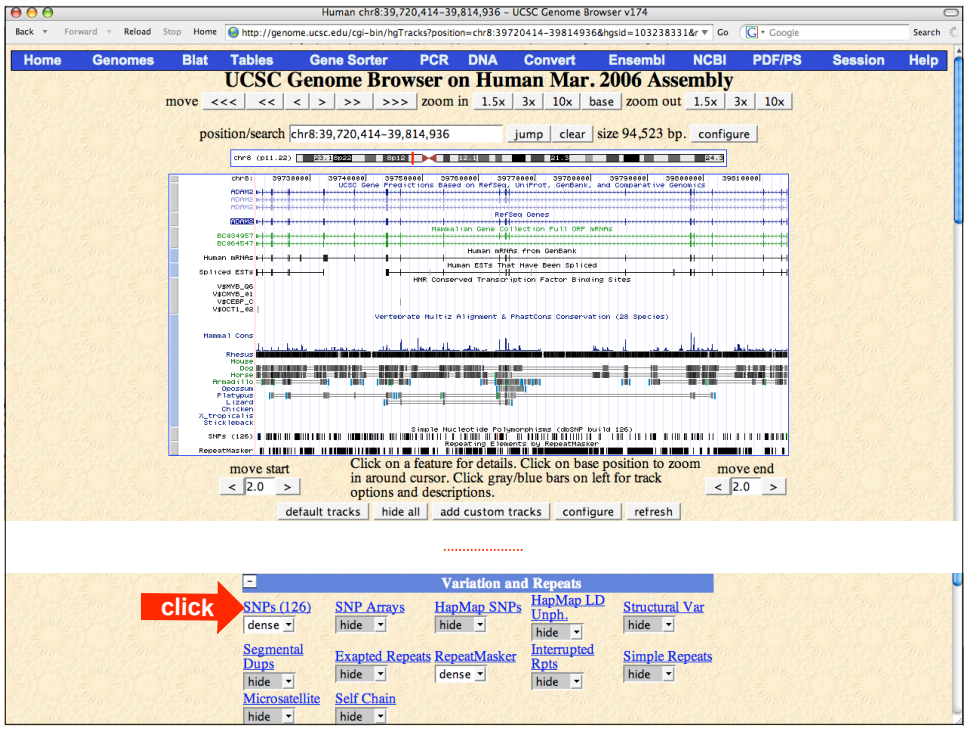

UCSC TFBS Track
Human chr8:39,625,891-39,909,459 - UCSC Genome Browser v174
Home
Genomes
Blat
Tables
Gene Sorter
PCR
DNA
Convert
Ensembl
NCBI
PDF/PS
Session
Help
UCSC Genome Browser on Human Mar. 2006 Assembly
position/search chr8:39,625,891-39,909,459
jump
clear
size 283,569 bp.
configure
click x 3
move start
move end
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
Map Contigs
Assembly
Gap
Coverage
BAC End Pairs

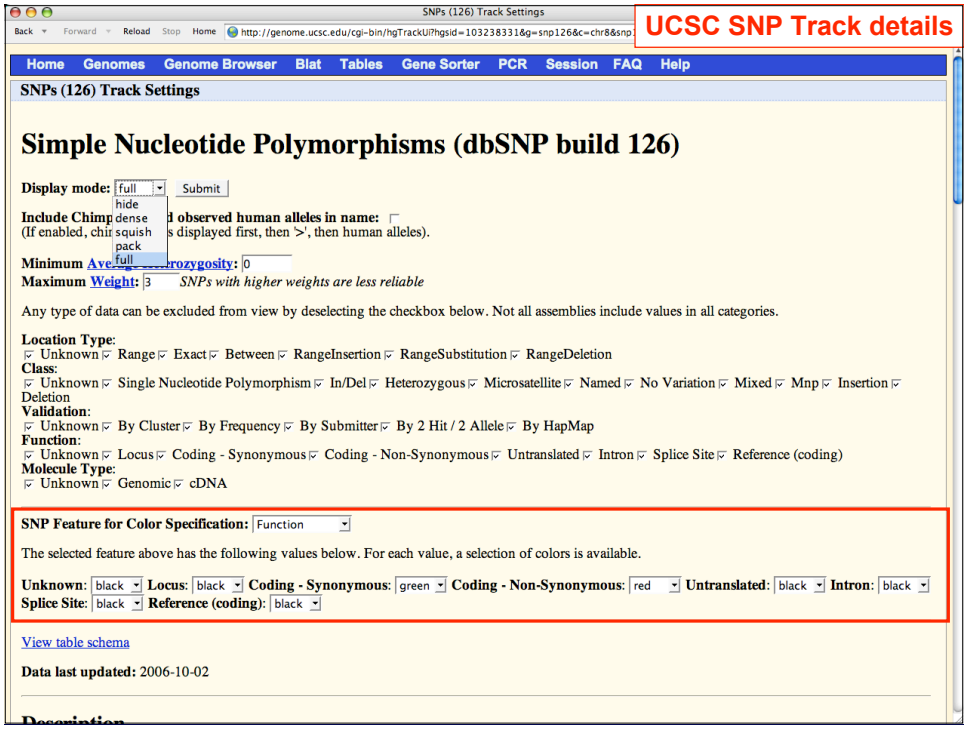

Human chr8:39,884,779-39,895,280 - UCSC Genome Browser v174
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help
UCSC Genome Browser on Human Mar. 2006 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chr8:39,884,779-39,895,280 jump clear size 10,502 bp. configure
UTR CDS
click
move start
move end
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
Map Contigs
Assembly
Gap
Coverage
BAC End Pairs
Fosmid End Pairs
GC Percent
Short Match
Restr Enzymes
Phenotype and Disease Associations
GAD
Case Control
NIMH Bipolar
RGD Human QTL
RGD Rat QTL

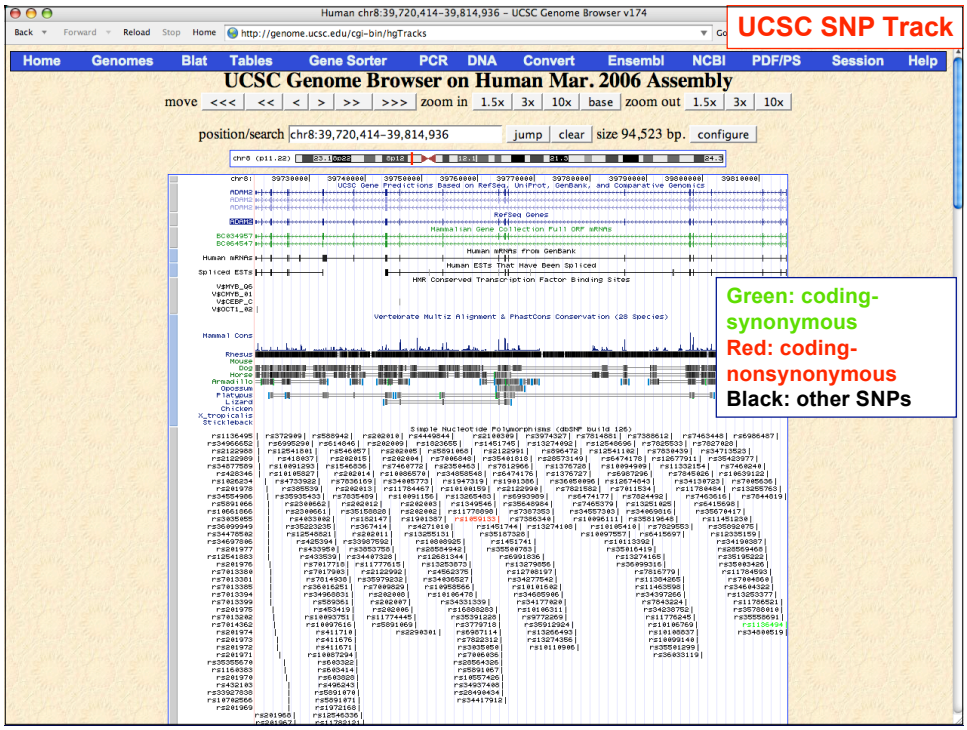

UCSC TFBS Track details
Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help
HMR Conserved Transcription Factor Binding Sites (V$IRF2_01)
Transcription Factor Binding Site information:
Item: V$IRF2_01
Transfac matrix link: M00063
Score: 891
zScore: 3.04
Strand: -
Position: chr8:39889441-39889453
Band: 8p11.22
Genomic Size: 13
View DNA for this feature
Transcription Factors known to bind to this site:
Factor: IRF-2
Species: mouse
SwissProt ID: P23906
View table schema
Data last updated: 2007-07-17
Description
This track contains the location and score of transcription factor binding sites conserved in the human/mouse/rat alignment. A binding site is considered to be conserved across the alignment if its score meets the threshold score for its binding matrix in all 3 species. The score and threshold are computed with the Transfac Matrix Database (v7.0) created by Biobase. The data are purely computational, and as such not all binding sites listed here are biologically functional binding sites.
In the graphical display, each box represents one conserved putative tfbs. Clicking on a box brings up detailed information on the binding site, namely its Transfac I.D., a link to its Transfac Matrix (free registration with Transfac required), its location in the human genome (chromosome, start, end, and strand), its length in bases, its raw score, and its Z score.
All binding factors that are known to bind to the particular binding matrix of the binding site are listed along with their species, SwissProt ID, and a link to that factor's page on the UCSC Protein Browser if such an entry exists.

UCSC

Change the color of items in a track

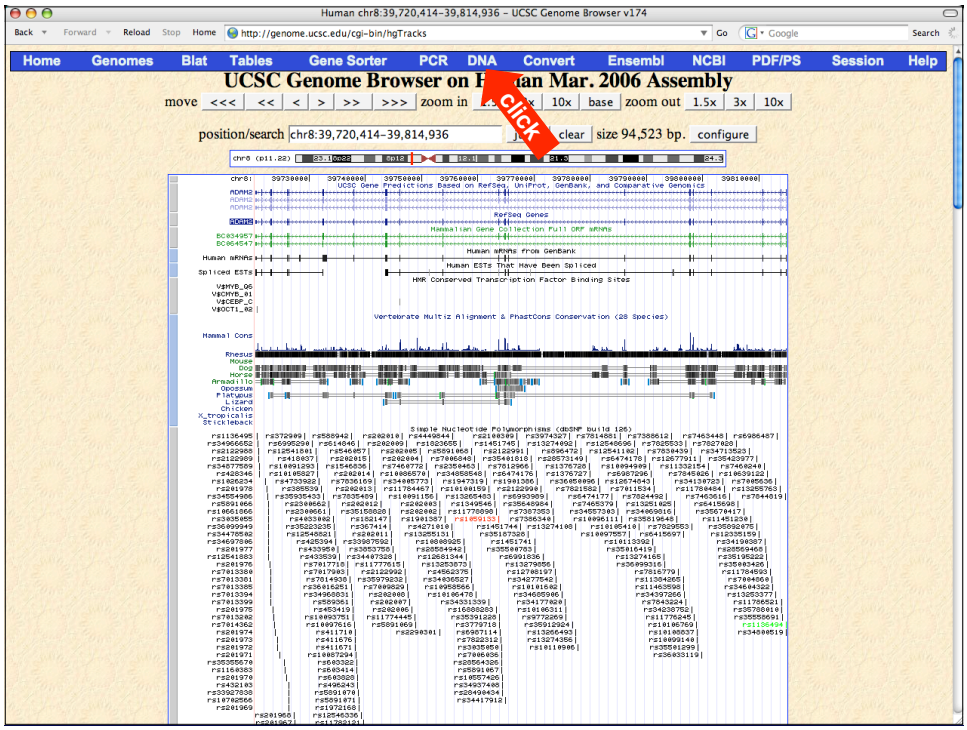


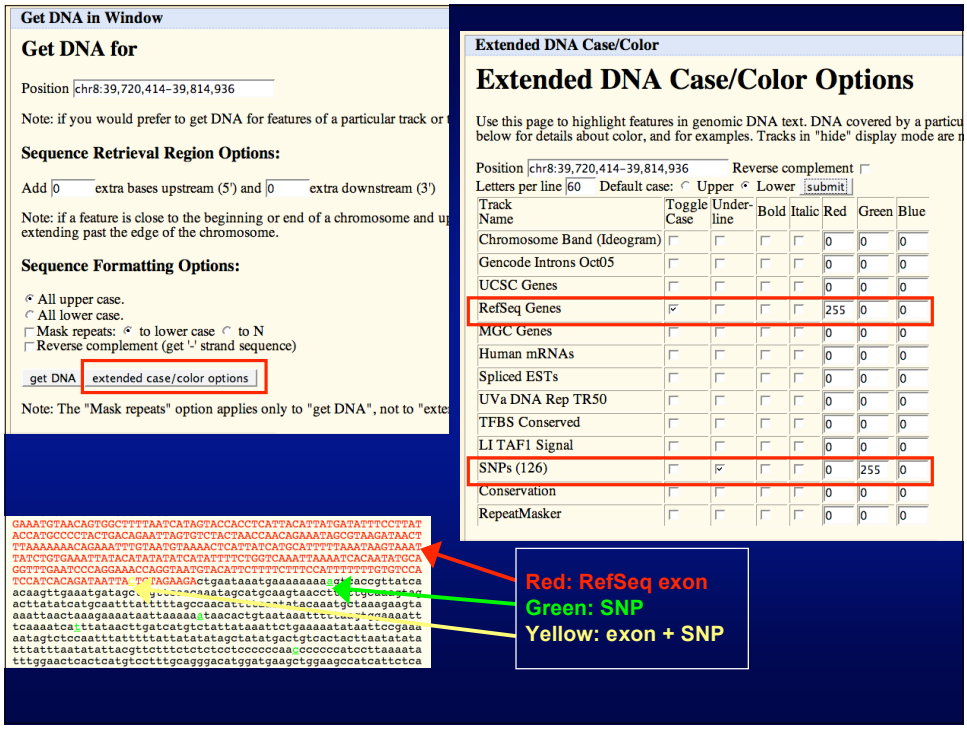

UCSC SNP Track details
SNPs (126) Track Settings
Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help
SNPs (126) Track Settings
Simple Nucleotide Polymorphisms (dbSNP build 126)
Display mode: full Submit
hide
dense
squish
pack
full
Minimum Weight: 0
Maximum Weight: 3 SNPs with higher weights are less reliable
Any type of data can be excluded from view by deselecting the checkbox below. Not all assemblies include values in all categories.
Location Type:
Unknown Range Exact Between RangeInsertion RangeSubstitution RangeDeletion
Class:
Unknown Single Nucleotide Polymorphism In/Del Heterozygous Microsatellite Named No Variation Mixed Mnp Insertion Deletion
Validation:
Unknown By Cluster By Frequency By Submitter By 2 Hit / 2 Allele By HapMap
Function:
Unknown Locus Coding - Synonymous Coding - Non-Synonymous Untranslated Intron Splice Site Reference (coding)
Molecule Type:
Unknown Genomic cDNA
SNP Feature for Color Specification: Function
The selected feature above has the following values below. For each value, a selection of colors is available.
Unknown: black Locus: black Coding - Synonymous: green Coding - Non-Synonymous: red Untranslated: black Intron: black
Splice Site: black Reference (coding): black
View table schema
Data last updated: 2006-10-02


UCSC SNP Track
Human chr8:39,720,414-39,814,936 - UCSC Genome Browser v174
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help
UCSC Genome Browser on Human Mar. 2006 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chr8:39,720,414-39,814,936 jump clear size 94,523 bp. configure
Green: coding-synonymous
Red: coding-nonsynonymous
Black: other SNPs

UCSC

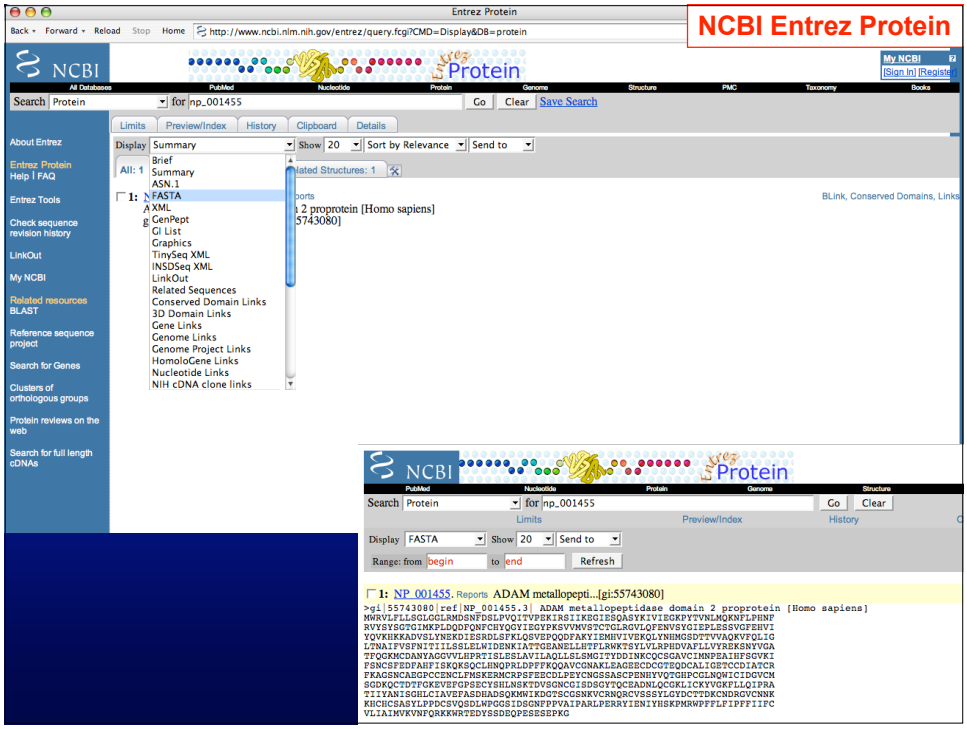
View features by changing the color of the genome sequence

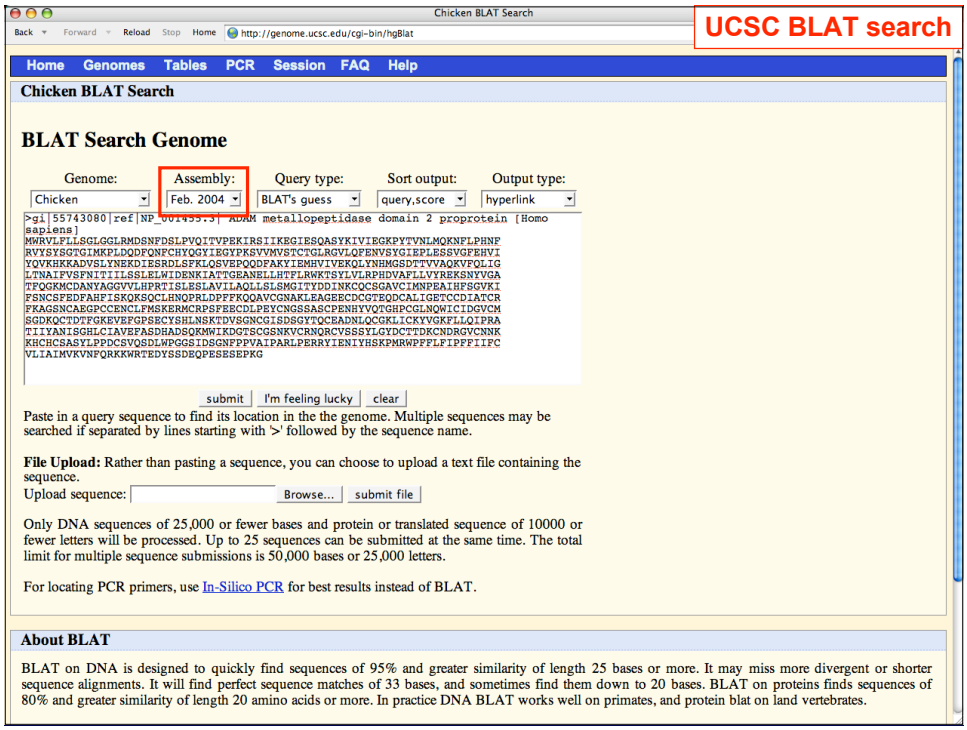


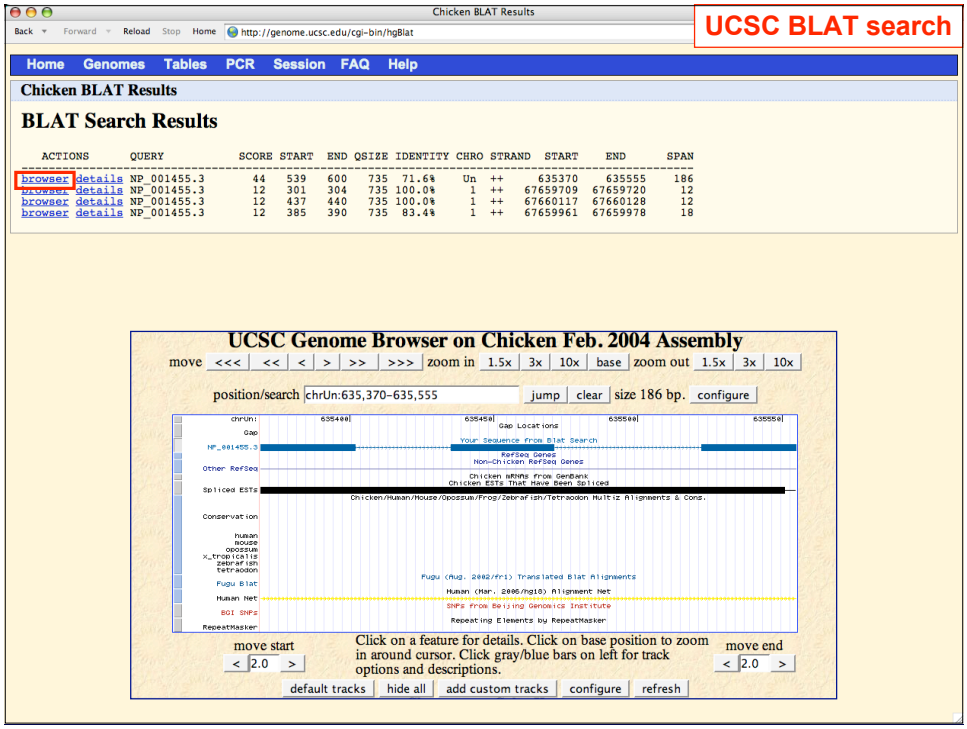


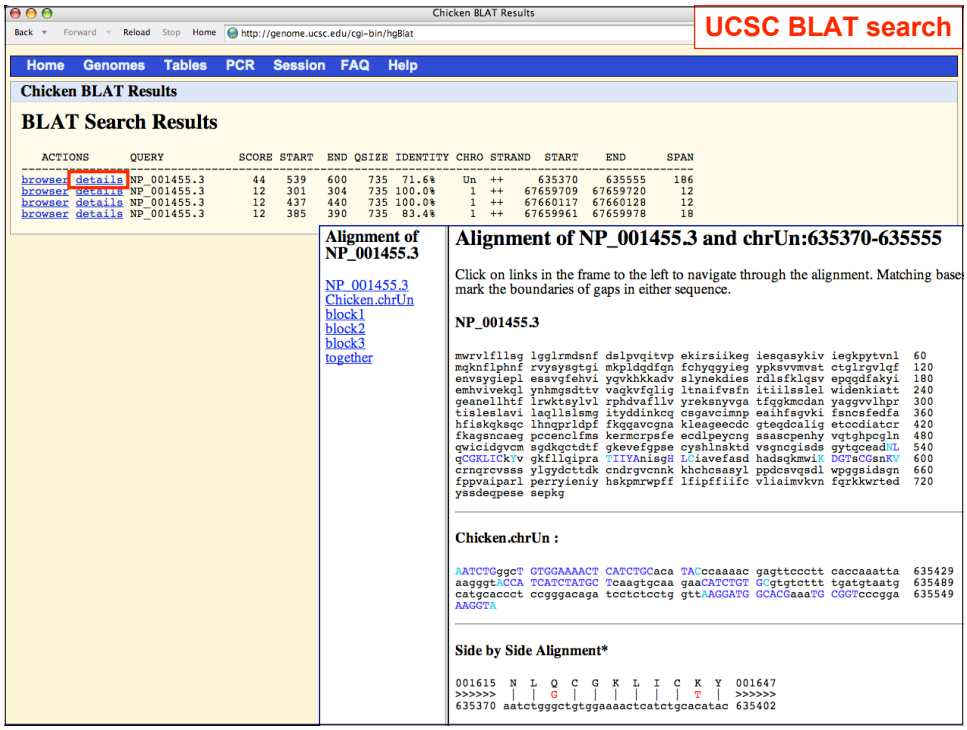
UCSC

Find a chicken homolog of a human protein


NCBI Entrez Protein
Entrez Protein
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Display&DB=protein
NCBI
Protein
Search Protein for np_001455
Go Clear Save Search
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Protein Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI
Related resources BLAST
Reference sequence project
Search for Genes
Clusters of orthologous groups
Protein reviews on the web
Search for full length cDNAs
Display Summary Show 20 Sort by Relevance Send to
Brief
Summary
ASN.1
FASTA
XML
GenPept
GI List
Graphics
TinySeq XML
INSDSeq XML
LinkOut
Related Sequences
Conserved Domain Links
3D Domain Links
Gene Links
Genome Links
Genome Project Links
HomoloGene Links
Nucleotide Links
NIH cDNA clone links
BLink, Conserved Domains, Links
Search Protein for np_001455
Limits Preview/Index History
Display FASTA Show 20 Send to
Range: from begin to end Refresh
1: NP_001455. Reports ADAM metallopepti...[gi:55743080]
>gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKIVIEGKPYTVNLMQKNFLPHNF
RVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPLESSVGFEHVI
YQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDFAKYIEMHVIVEKQLVNHMGSDTTVVAQKVFQLIQ
LTNAIFVSFNITILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLRPHDVAFLLVYREKSNYVGA
TFQGKMCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAIHFSGVKI
FSNCSFEDFAHFISKQKSQCLHNQPRLDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCM
SGDKQCTDTFGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYTQCEADNLQCGKLICKYVGKFLLQIPRA
TIIYANISGHLCIAVEFASDHADSQPMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTDKCNDRGVCNNK
KHCHCSASYLPPDCSVQSDLWPGGSIDSGNFPPVAIPARLPERRYIENIYHSKPMRWPFFLFIPFFIIFC
VLIAIMVKVNFQRKKWRTEDYSSDEQPESESEPKG

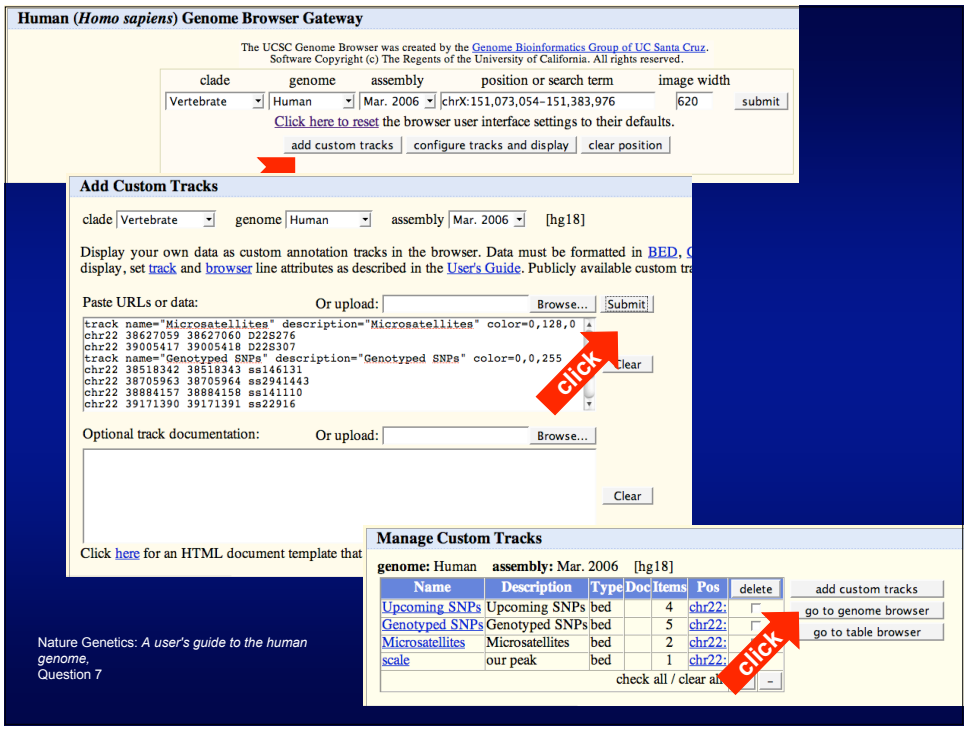

UCSC BLAT search
Chicken BLAT Search
http://genome.ucsc.edu/cgi-bin/hgBlat
Home Genomes Tables PCR Session FAQ Help
Chicken BLAT Search
BLAT Search Genome
Genome: Chicken
Assembly: Feb. 2004
Query type: BLAT's guess
Sort output: query,score
Output type: hyperlink
>gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKIVIEGKPYTVNLMQKNFLPHNF
RVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPLESSVGFEHVI
YQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDFAKYIEMHVIVEKQLVNHMGSDTTVVAQKVFQLIQ
LTNAIFVSFNITILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLRPHDVAFLLVYREKSNYVGA
TFQGKMCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAIHFSGVKI
FSNCSFEDFAHFISKQKSQCLHNQPRLDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCM
SGDKQCTDTFGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYTQCEADNLQCGKLICKYVGKFLLQIPRA
TIIYANISGHLCIAVEFASDHADSQPMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTDKCNDRGVCNNK
KHCHCSASYLPPDCSVQSDLWPGGSIDSGNFPPVAIPARLPERRYIENIYHSKPMRWPFFLFIPFFIIFC
VLIAIMVKVNFQRKKWRTEDYSSDEQPESESEPKG
submit I'm feeling lucky clear
Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.
File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.
Upload sequence: Browse... submit file
Only DNA sequences of 25,000 or fewer bases and protein or translated sequence of 10000 or fewer letters will be processed. Up to 25 sequences can be submitted at the same time. The total limit for multiple sequence submissions is 50,000 bases or 25,000 letters.
For locating PCR primers, use In-Silico PCR for best results instead of BLAT.
About BLAT
BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 25 bases or more. It may miss more divergent or shorter sequence alignments. It will find perfect sequence matches of 33 bases, and sometimes find them down to 20 bases. BLAT on proteins finds sequences of 80% and greater similarity of length 20 amino acids or more. In practice DNA BLAT works well on primates, and protein blat on land vertebrates.

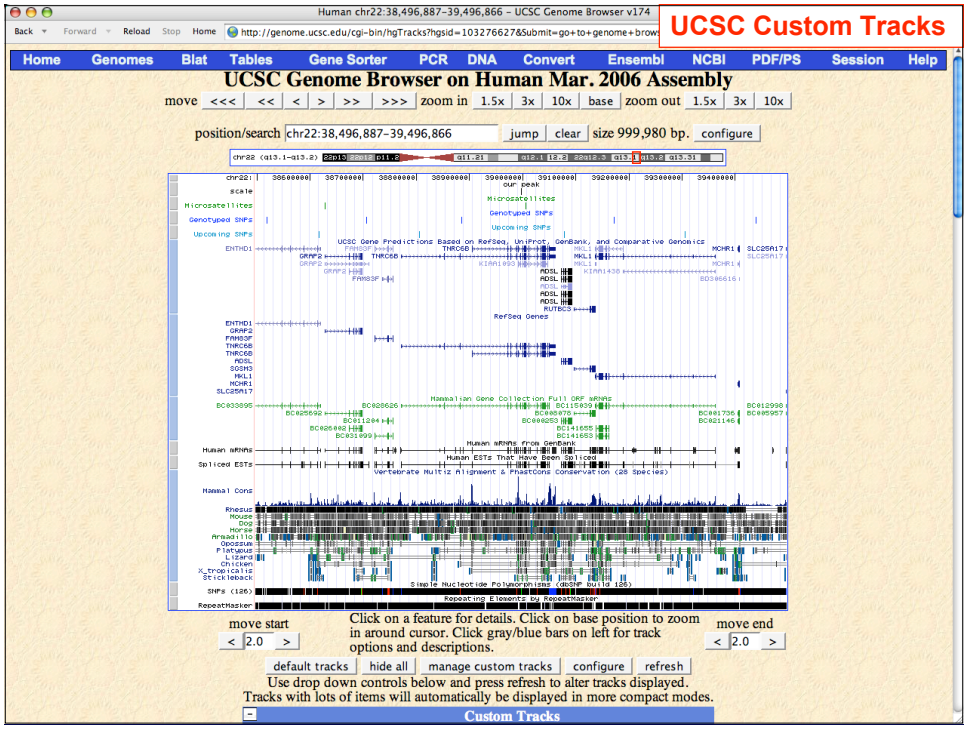

Chicken BLAT Results
UCSC BLAT search
Home Genomes Tables PCR Session FAQ Help
Chicken BLAT Results
BLAT Search Results
ACTIONS QUERY SCORE START END QSIZE IDENTITY CHRO STRAND START END SPAN
browser details NP_001455.3 44 539 600 735 71.6% Un ++ 635370 635555 186
browser details NP_001455.3 12 301 304 735 100.0% 1 ++ 67659709 67659720 12
browser details NP_001455.3 12 437 440 735 100.0% 1 ++ 67660117 67660128 12
browser details NP_001455.3 12 385 390 735 83.4% 1 ++ 67659961 67659978 18
UCSC Genome Browser on Chicken Feb. 2004 Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chrUn:635,370-635,555 jump clear size 186 bp. configure
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
move end
default tracks hide all add custom tracks configure refresh


Chicken BLAT Results
UCSC BLAT search
Home Genomes Tables PCR Session FAQ Help
Chicken BLAT Results
BLAT Search Results
ACTIONS QUERY SCORE START END QSIZE IDENTITY CHRO STRAND START END SPAN
browser details NP_001455.3 44 539 600 735 71.6% Un ++ 635370 635555 186
browser details NP_001455.3 12 301 304 735 100.0% 1 ++ 67659709 67659720 12
browser details NP_001455.3 12 437 440 735 100.0% 1 ++ 67660117 67660128 12
browser details NP_001455.3 12 385 390 735 83.4% 1 ++ 67659961 67659978 18
Alignment of NP_001455.3
NP_001455.3
Chicken.chrUn
block1
block2
block3
together
Alignment of NP_001455.3 and chrUn:635370-635555
Click on links in the frame to the left to navigate through the alignment. Matching bases mark the boundaries of gaps in either sequence.
NP_001455.3
Chicken.chrUn :
Side by Side Alignment*

UCSC

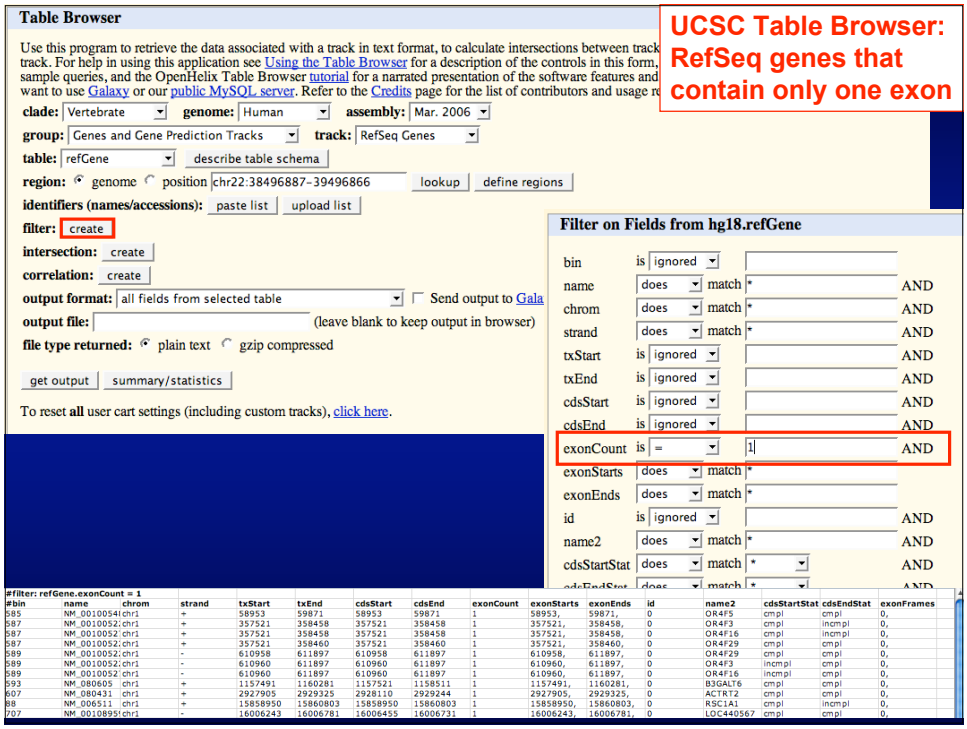
Add your own custom tracks





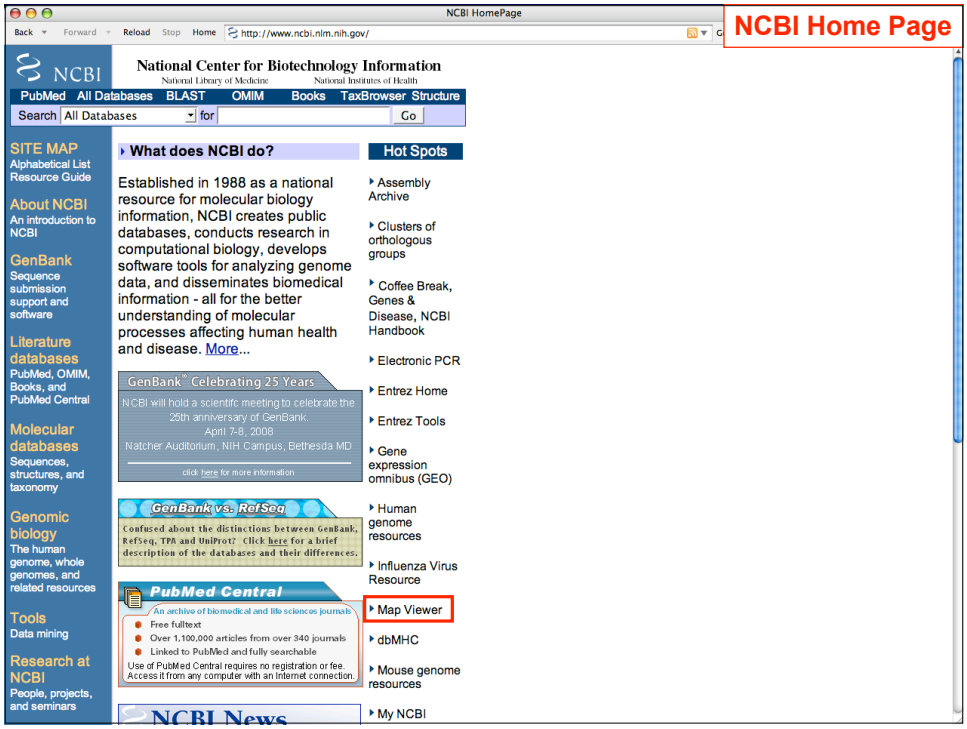
UCSC Table Browser

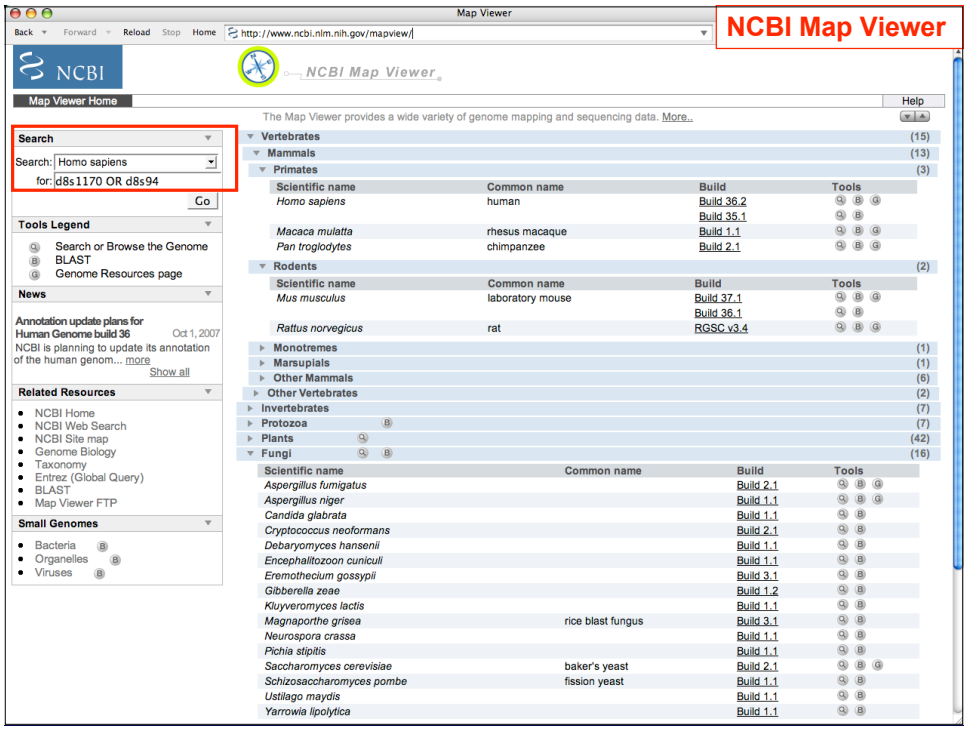
- Download track in text format
- Retrieve DNA sequence covered by a track
- Calculate intersections between tracks and view in the Genome Browser. For example:
  - Show all RefSeq genes that contain only one exon
  - Show transcription factor binding sites that overlap (intersect) with a SNP



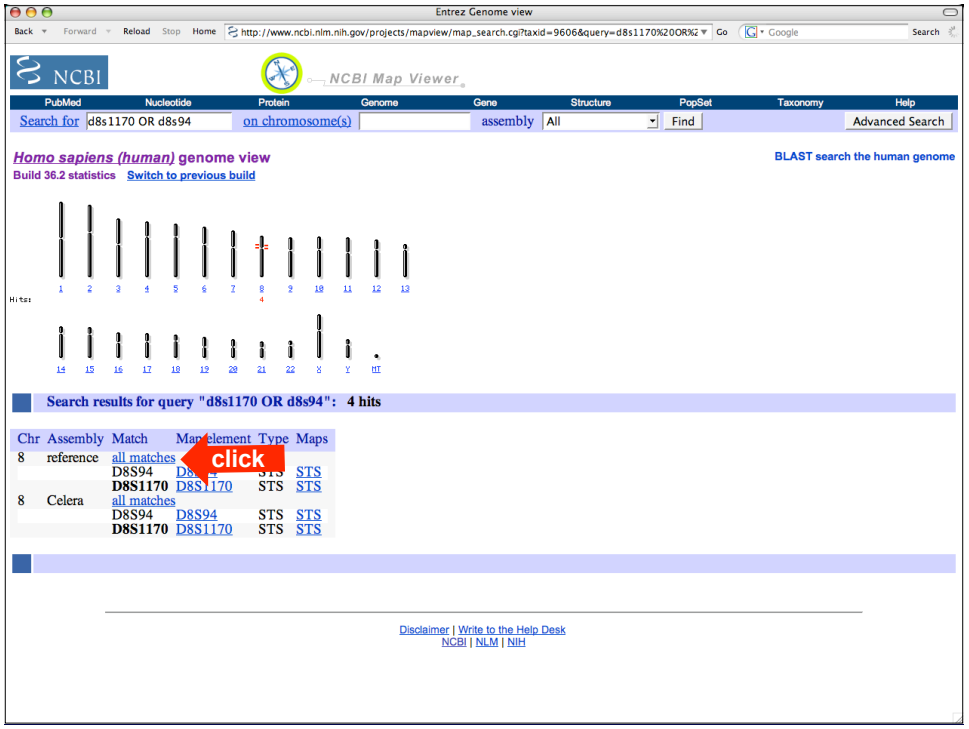
NCBI

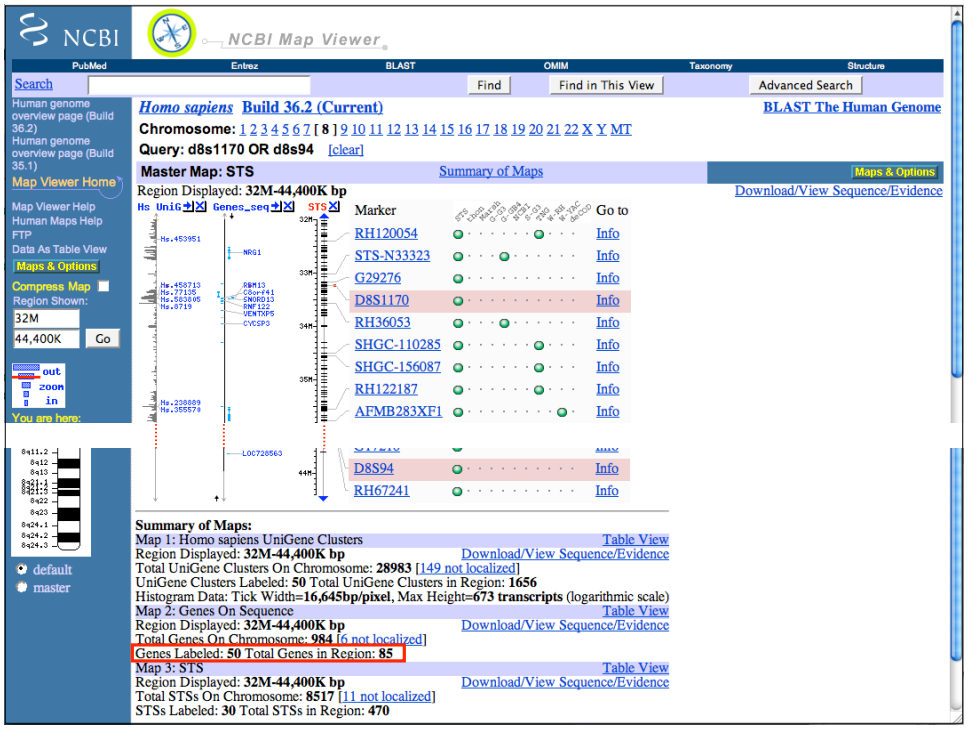
View a genomic region between two STS markers





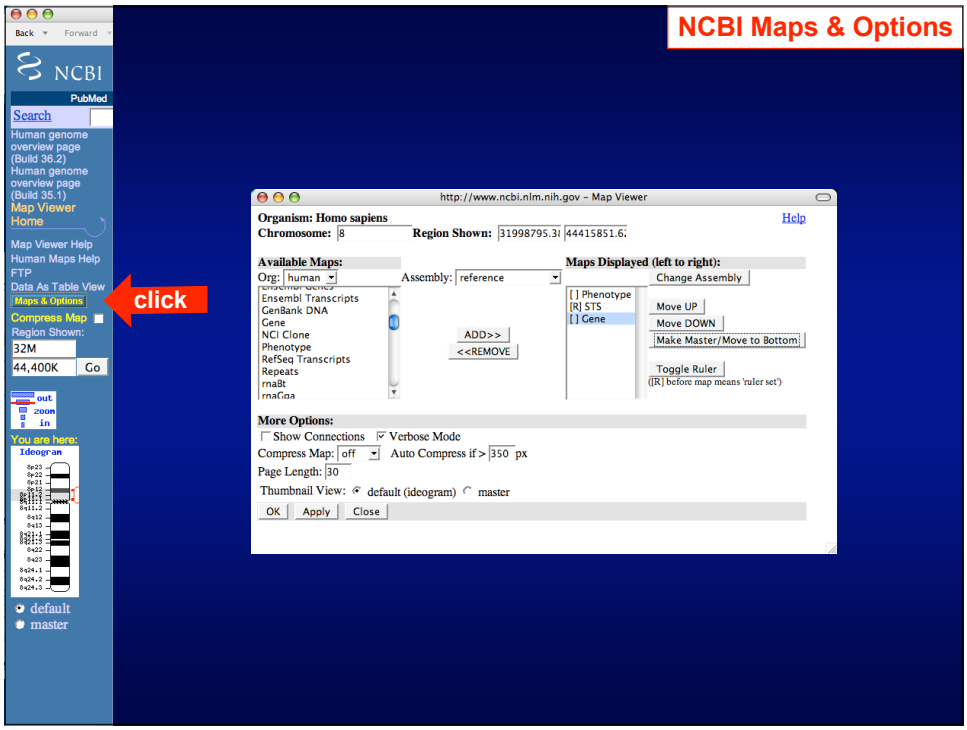
NCBI Map Viewer

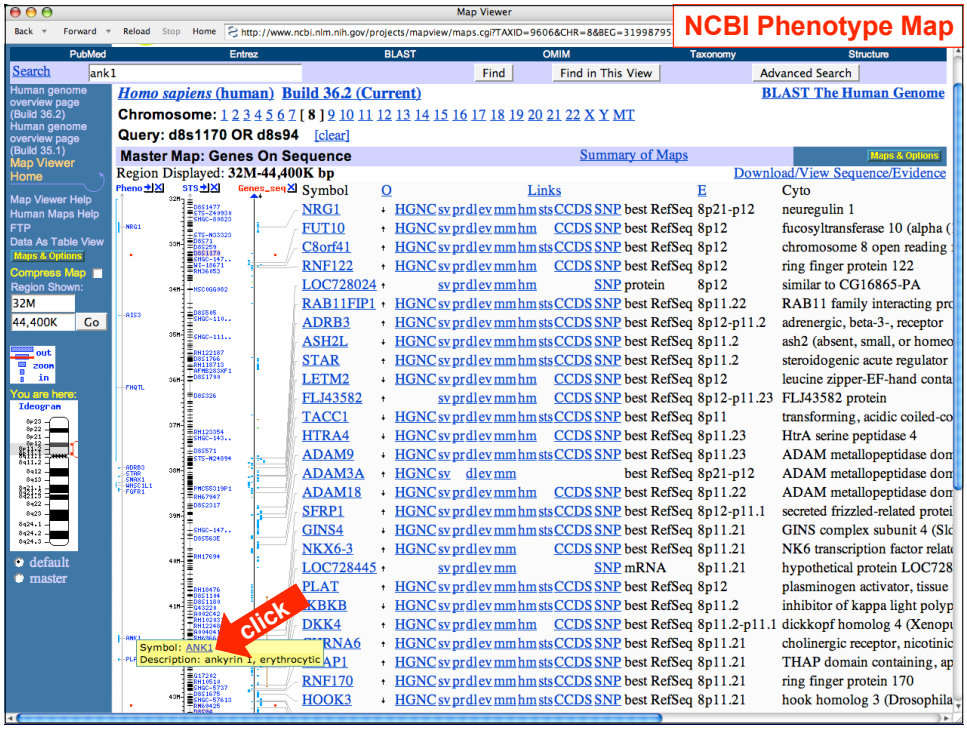

Entrez Genome view
NCBI
NCBI Map Viewer
PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help
Search for d8s1170 OR d8s94 on chromosome(s) assembly All Find Advanced Search
Homo sapiens (human) genome view
Build 36.2 statistics Switch to previous build
BLAST search the human genome
Search results for query "d8s1170 OR d8s94": 4 hits
Chr Assembly Match Map element Type Maps
8 reference all matches
D8S94 D8S94 STS STS
D8S1170 D8S1170 STS STS
8 Celera all matches
D8S94 D8S94 STS STS
D8S1170 D8S1170 STS STS
click
Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

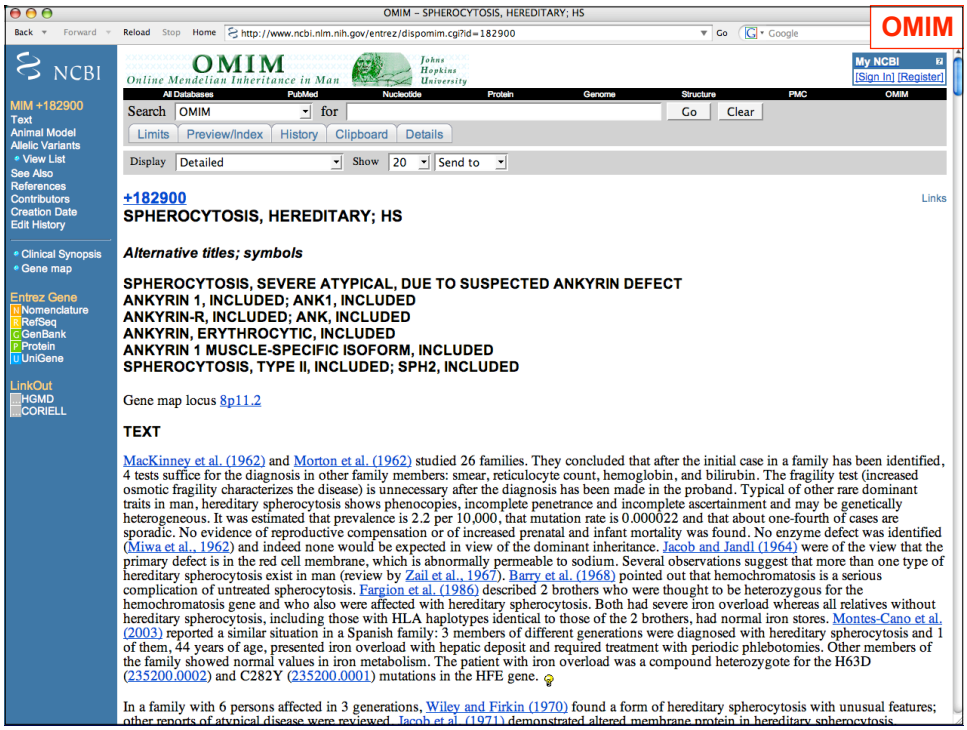

NCBI
NCBI Map Viewer
PubMed Entrez BLAST OMIM Taxonomy Structure
Search Find Find in This View Advanced Search
Homo sapiens Build 36.2 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: d8s1170 OR d8s94 [clear]
Master Map: STS
Summary of Maps
Maps & Options
Download/View Sequence/Evidence
Region Displayed: 32M-44,400K bp
Marker
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
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
Compress Map
Region Shown:
32M
44,400K
Go
out
zoom
in
You are here:
default
master
Summary of Maps:
Map 1: Homo sapiens UniGene Clusters
Table View
Region Displayed: 32M-44,400K bp
Download/View Sequence/Evidence
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
Table View
Region Displayed: 32M-44,400K bp
Download/View Sequence/Evidence
Total Genes On Chromosome: 984 [6 not localized]
Genes Labeled: 50 Total Genes in Region: 85
Map 3: STS
Table View
Region Displayed: 32M-44,400K bp
Download/View Sequence/Evidence
Total STSs On Chromosome: 8517 [11 not localized]
STSs Labeled: 30 Total STSs in Region: 470

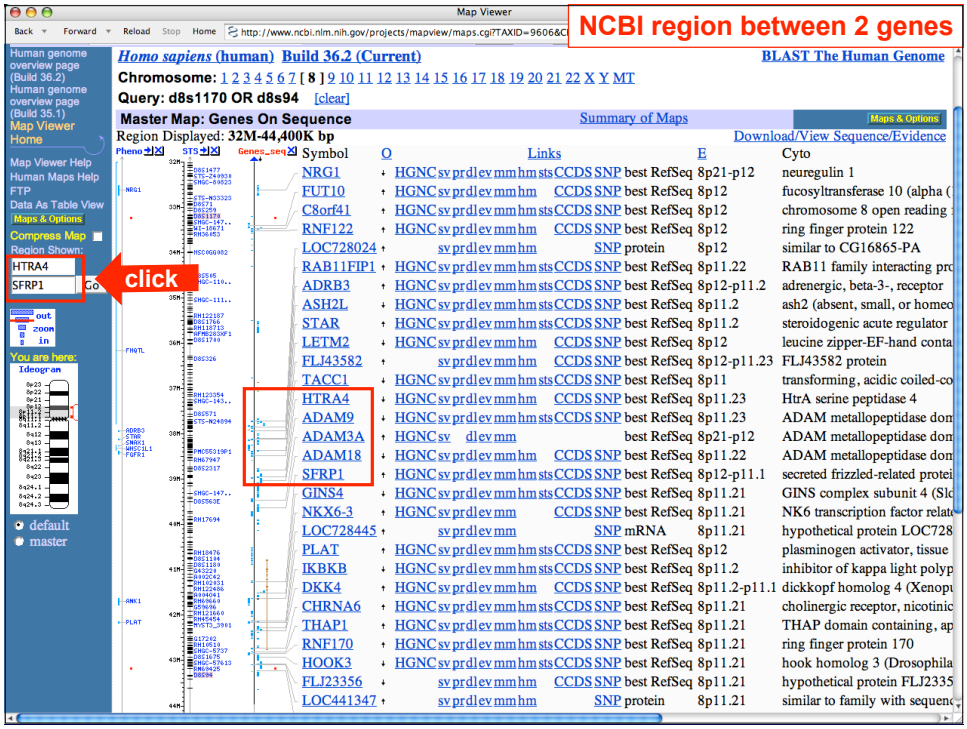
NCBI

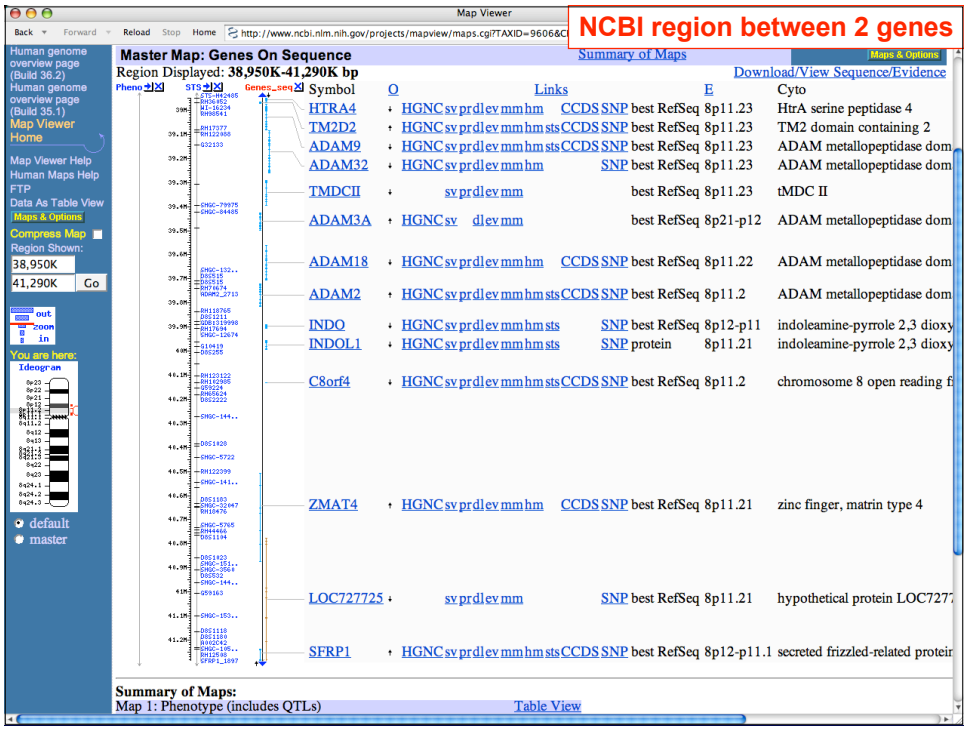
Change the maps displayed on the Map Viewer




NCBI Phenotype Map
Map Viewer
Homo sapiens (human) Build 36.2 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: d8s1170 OR d8s94
Master Map: Genes On Sequence
Summary of Maps
Region Displayed: 32M-44,400K bp
Download/View Sequence/Evidence
click
Symbol: ANK1
Description: ankyrin 1, erythrocytic

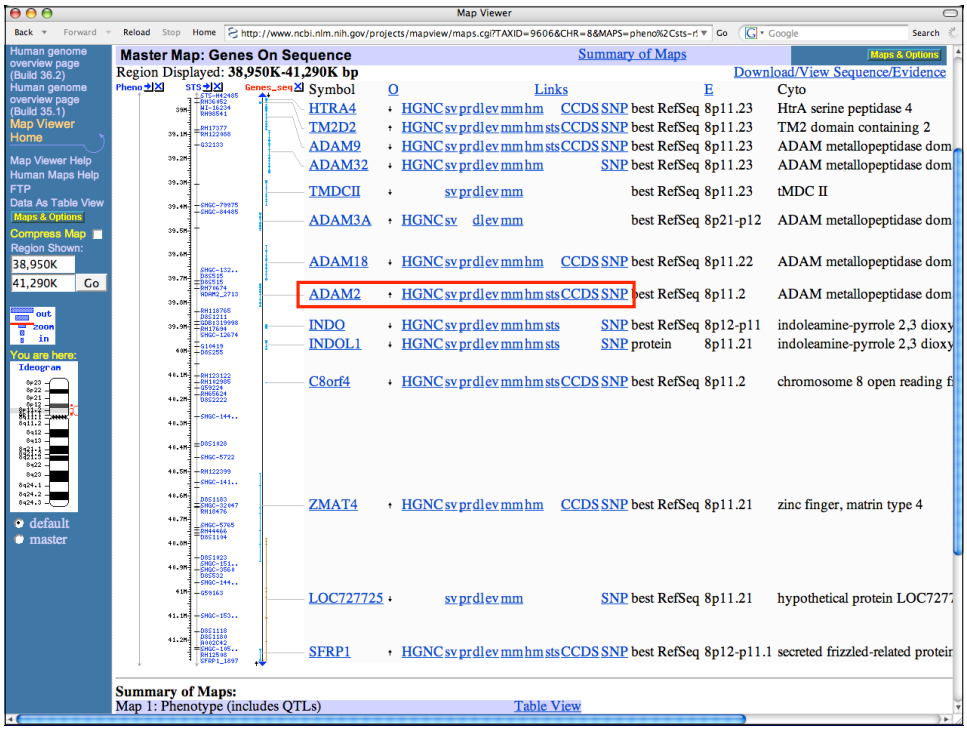

OMIM
OMIM – SPHEROCYTOSIS, HEREDITARY; HS
+182900
SPHEROCYTOSIS, HEREDITARY; HS
Alternative titles; symbols
SPHEROCYTOSIS, SEVERE ATYPICAL, DUE TO SUSPECTED ANKYRIN DEFECT
ANKYRIN 1, INCLUDED; ANK1, INCLUDED
ANKYRIN-R, INCLUDED; ANK, INCLUDED
ANKYRIN, ERYTHROCYTIC, INCLUDED
ANKYRIN 1 MUSCLE-SPECIFIC ISOFORM, INCLUDED
SPHEROCYTOSIS, TYPE II, INCLUDED; SPH2, INCLUDED
Gene map locus 8p11.2
TEXT

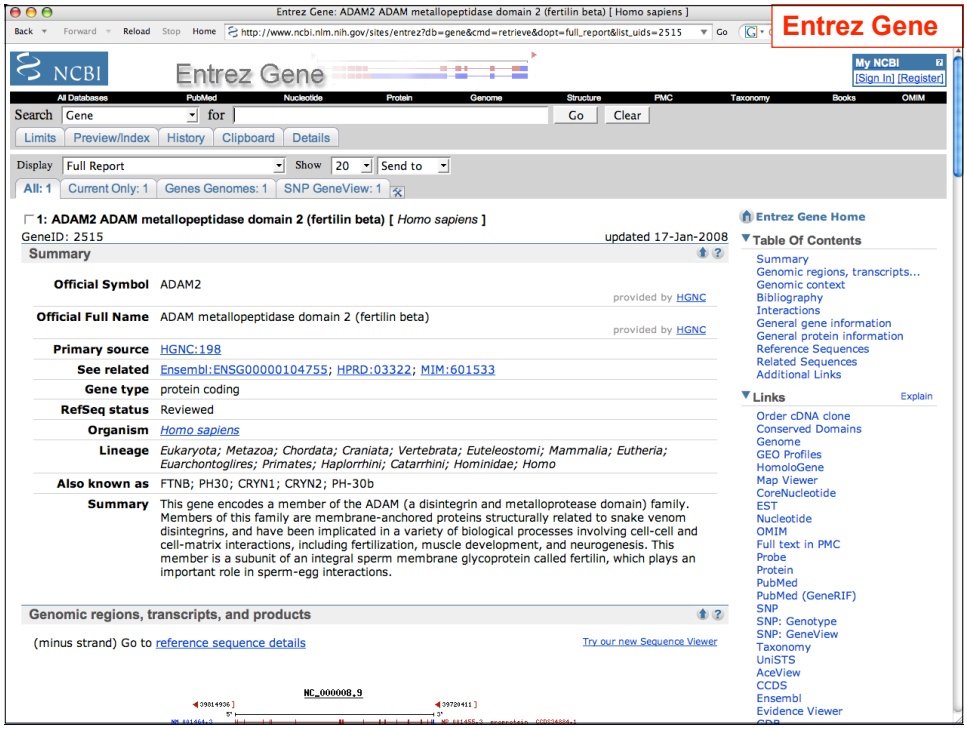

NCBI region between 2 genes
Map Viewer
Homo sapiens (human) Build 36.2 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: d8s1170 OR d8s94 [clear]
Master Map: Genes On Sequence
Summary of Maps
Region Displayed: 32M-44,400K bp
Download/View Sequence/Evidence
HTRA4
SFRP1
Go
click
Symbol
Links
Cyto
NRG1 8p21-p12 neuregulin 1
FUT10 8p12 fucosyltransferase 10 (alpha (
C8orf41 8p12 chromosome 8 open reading
RNF122 8p12 ring finger protein 122
LOC728024 8p12 similar to CG16865-PA
RAB11FIP1 8p11.22 RAB11 family interacting pro
ADRB3 8p12-p11.2 adrenergic, beta-3-, receptor
ASH2L 8p11.2 ash2 (absent, small, or homeo
STAR 8p11.2 steroidogenic acute regulator
LETM2 8p12 leucine zipper-EF-hand conta
FLJ43582 8p12-p11.23 FLJ43582 protein
TACC1 8p11 transforming, acidic coiled-co
HTRA4 8p11.23 HtrA serine peptidase 4
ADAM9 8p11.23 ADAM metallopeptidase dom
ADAM3A 8p21-p12 ADAM metallopeptidase dom
ADAM18 8p11.22 ADAM metallopeptidase dom
SFRP1 8p12-p11.1 secreted frizzled-related protei
GINS4 8p11.21 GINS complex subunit 4 (Sl
NKX6-3 8p11.21 NK6 transcription factor relat
LOC728445 8p11.21 hypothetical protein LOC728
PLAT 8p12 plasminogen activator, tissue
IKBKB 8p11.2 inhibitor of kappa light polyp
DKK4 8p11.2-p11.1 dickkopf homolog 4 (Xenop
CHRNA6 8p11.21 cholinergic receptor, nicotinic
THAP1 8p11.21 THAP domain containing, ap
RNF170 8p11.21 ring finger protein 170
HOOK3 8p11.21 hook homolog 3 (Drosophila
FLJ23356 8p11.21 hypothetical protein FLJ2335
LOC441347 8p11.21 similar to family with sequen

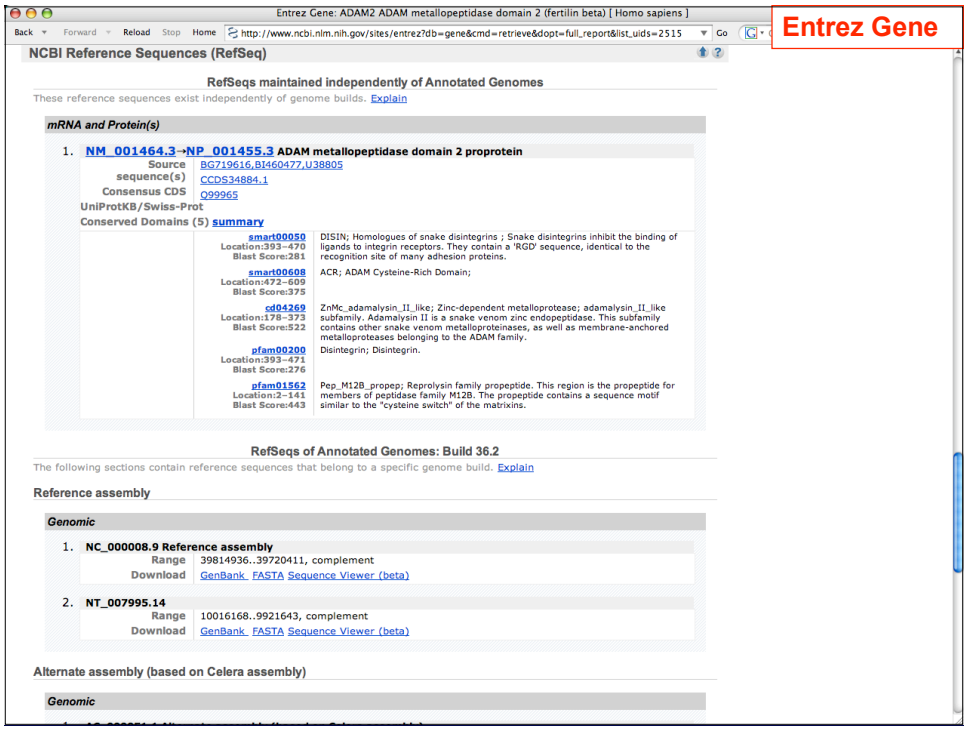

NCBI region between 2 genes
Map Viewer
Master Map: Genes On Sequence
Summary of Maps
Region Displayed: 38,950K-41,290K bp
Download/View Sequence/Evidence
38,950K
41,290K
Go
Symbol
Links
Cyto
HTRA4 8p11.23 HtrA serine peptidase 4
TM2D2 8p11.23 TM2 domain containing 2
ADAM9 8p11.23 ADAM metallopeptidase dom
ADAM32 8p11.23 ADAM metallopeptidase dom
TMDCII 8p11.23 tMDC II
ADAM3A 8p21-p12 ADAM metallopeptidase dom
ADAM18 8p11.22 ADAM metallopeptidase dom
ADAM2 8p11.2 ADAM metallopeptidase dom
INDO 8p12-p11 indoleamine-pyrrole 2,3 dioxy
INDOL1 8p11.21 indoleamine-pyrrole 2,3 dioxy
C8orf4 8p11.2 chromosome 8 open reading f
ZMAT4 8p11.21 zinc finger, matrin type 4
LOC727725 8p11.21 hypothetical protein LOC727
SFRP1 8p12-p11.1 secreted frizzled-related protei
Summary of Maps:
Map 1: Phenotype (includes QTLs)
Table View

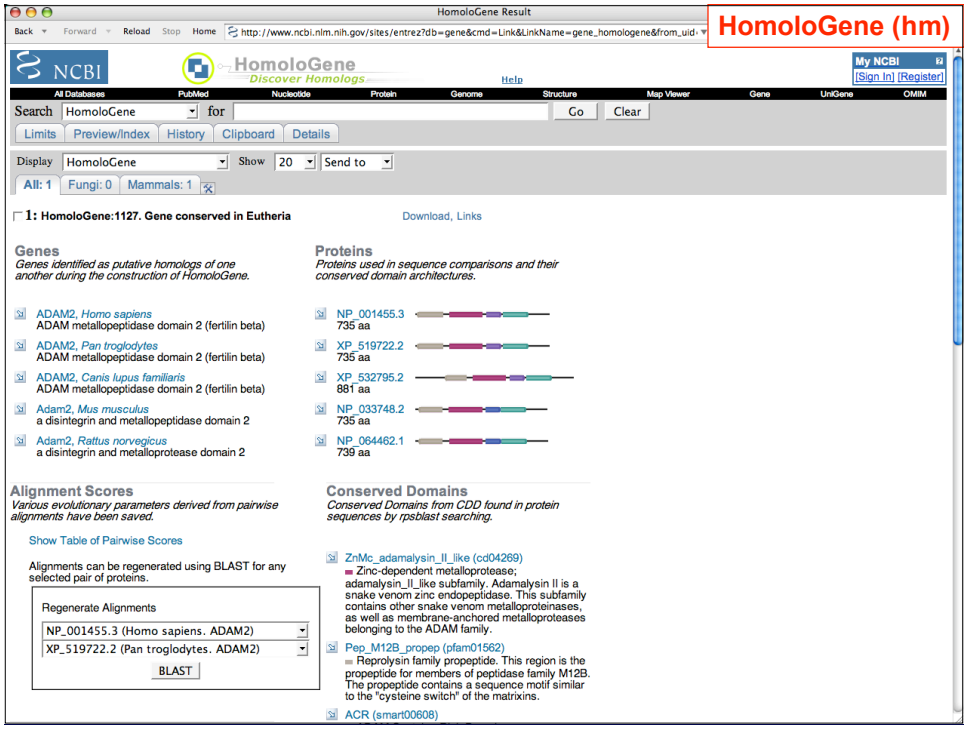
NCBI

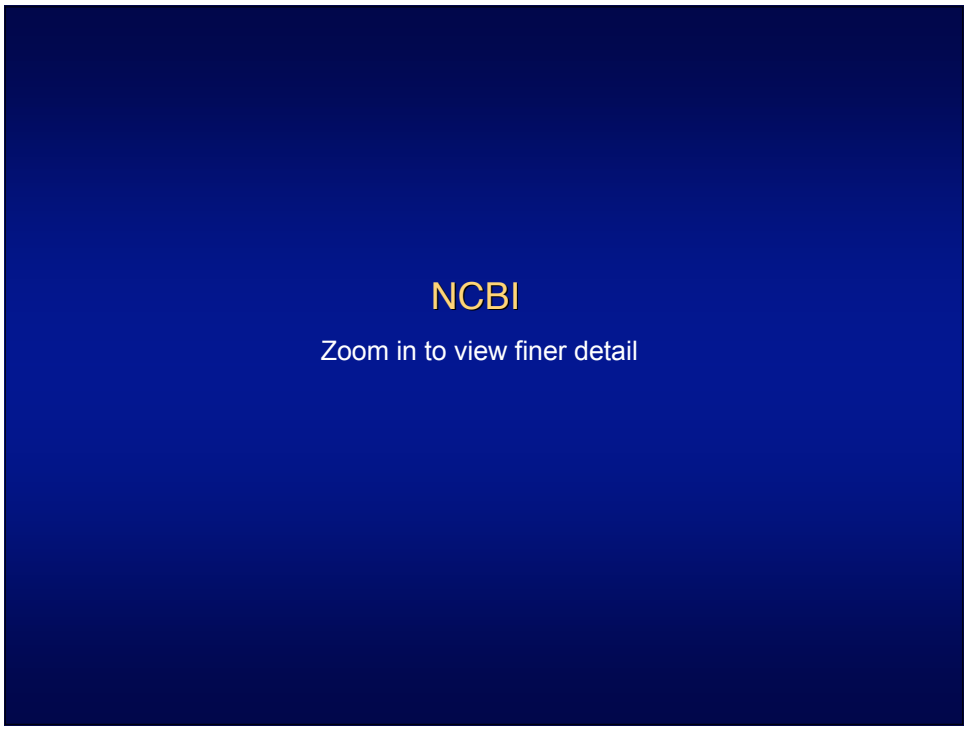
View additional information about a gene

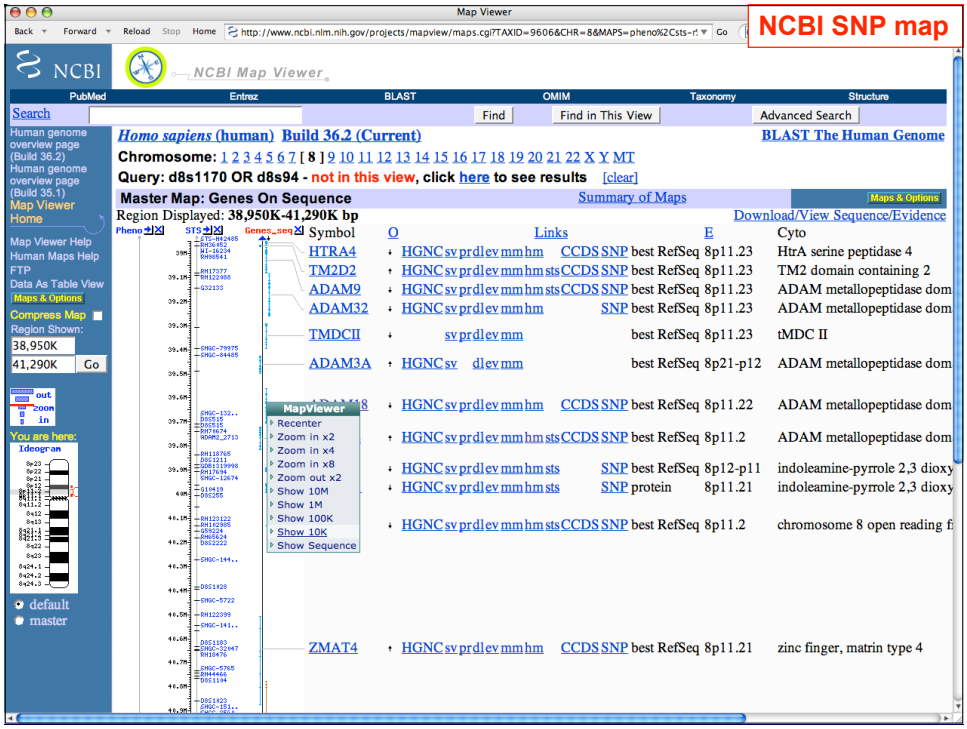


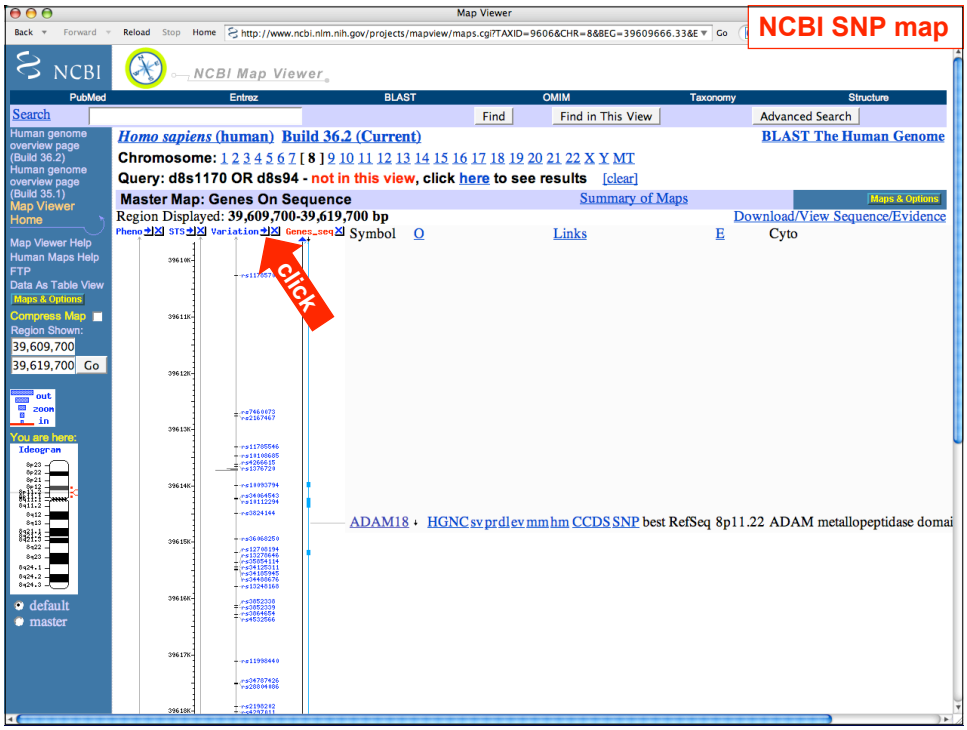

Entrez Gene
Entrez Gene: ADAM2 ADAM metallopeptidase domain 2 (fertilin beta) [Homo sapiens]
Entrez Gene
1: ADAM2 ADAM metallopeptidase domain 2 (fertilin beta) [Homo sapiens]
GeneID: 2515
updated 17-Jan-2008
Summary
Official Symbol ADAM2
Official Full Name ADAM metallopeptidase domain 2 (fertilin beta)
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
NC_000008.9

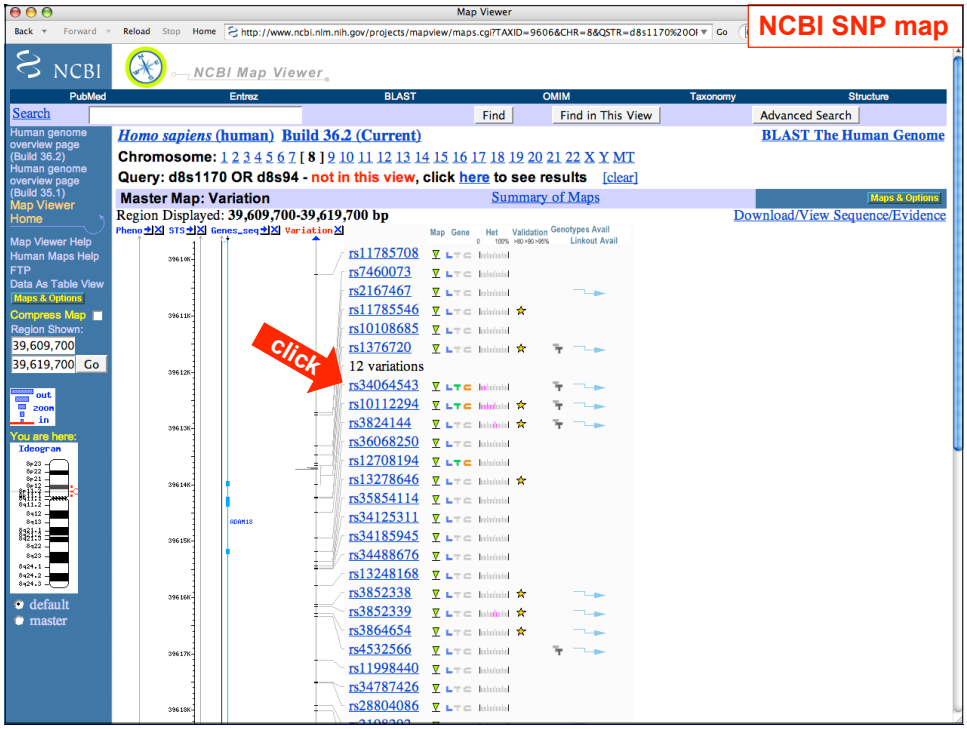

Entrez Gene
Entrez Gene: ADAM2 ADAM metallopeptidase domain 2 (fertilin beta) [Homo sapiens]
NCBI Reference Sequences (RefSeq)
RefSeqs maintained independently of Annotated Genomes
mRNA and Protein(s)
1. NM_001464.3→NP_001455.3 ADAM metallopeptidase domain 2 proprotein
RefSeqs of Annotated Genomes: Build 36.2
Reference assembly
Genomic
1. NC_000008.9 Reference assembly
Range 39814936..39720411, complement
2. NT_007995.14
Range 10016168..9921643, complement
Alternate assembly (based on Celera assembly)
Genomic

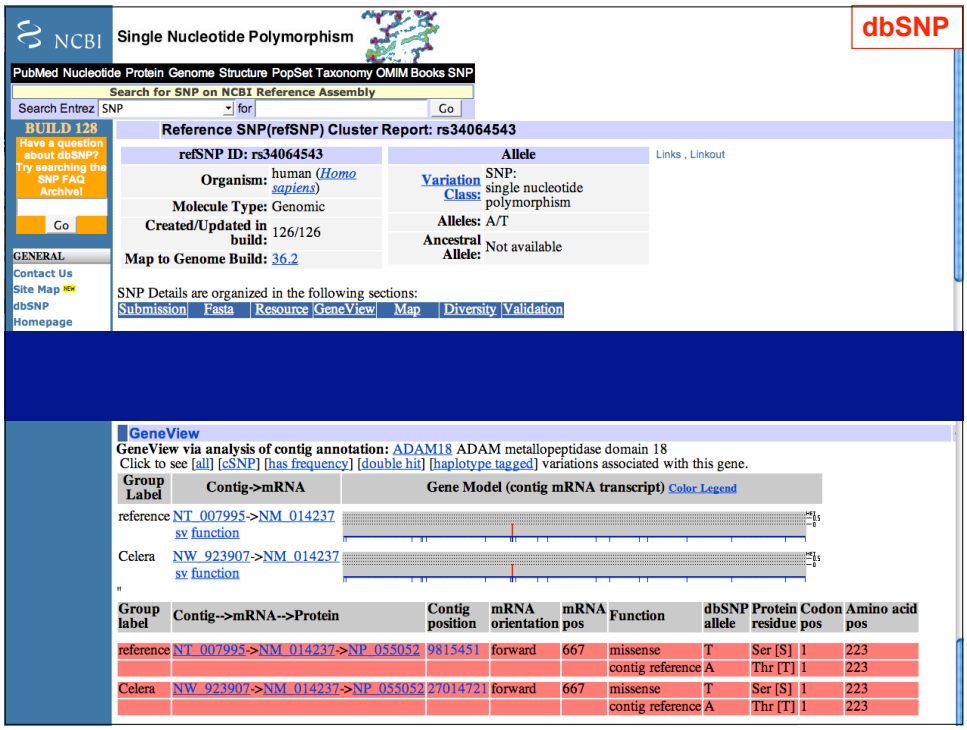

HomoloGene Result
HomoloGene (hm)
NCBI
HomoloGene
Discover Homologs
Help
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure Map Viewer Gene UniGene OMIM
Search HomoloGene for Go Clear
Limits Preview/Index History Clipboard Details
Display HomoloGene Show 20 Send to
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 metallopeptidase 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)
XP_519722.2 (Pan troglodytes. ADAM2)
BLAST
Conserved Domains
Conserved Domains from CDD found in protein sequences by rpsblast searching.
ZnMc_adamalysin_II_like (cd04269)
Zinc-dependent metalloprotease; adamalysin_II_like subfamily. Adamalysin 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)


NCBI
Zoom in to view finer detail


NCBI SNP map
Map Viewer
NCBI Map Viewer
Homo sapiens (human) Build 36.2 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: d8s1170 OR d8s94 - not in this view, click here to see results [clear]
Master Map: Genes On Sequence
Summary of Maps
Region Displayed: 38,950K-41,290K bp
Download/View Sequence/Evidence
HTRA4 HGNC sv prd lev mm hm CCDS SNP best RefSeq 8p11.23 HtrA serine peptidase 4
TM2D2 HGNC sv prd lev mm hm sts CCDS SNP best RefSeq 8p11.23 TM2 domain containing 2
ADAM9 HGNC sv prd lev mm hm sts CCDS SNP best RefSeq 8p11.23 ADAM metallopeptidase dom
ADAM32 HGNC sv prd lev mm hm SNP best RefSeq 8p11.23 ADAM metallopeptidase dom
TMDCII sv prd lev mm best RefSeq 8p11.23 tMDC II
ADAM3A HGNC sv dlev mm best RefSeq 8p21-p12 ADAM metallopeptidase dom
HGNC sv prd lev mm hm CCDS SNP best RefSeq 8p11.22 ADAM metallopeptidase dom
HGNC sv prd lev mm hm sts CCDS SNP best RefSeq 8p11.2 ADAM metallopeptidase dom
HGNC sv prd lev mm hm sts SNP best RefSeq 8p12-p11 indoleamine-pyrrole 2,3 dioxy
HGNC sv prd lev mm hm sts SNP protein 8p11.21 indoleamine-pyrrole 2,3 dioxy
HGNC sv prd lev mm hm sts CCDS SNP best RefSeq 8p11.2 chromosome 8 open reading f
ZMAT4 HGNC sv prd lev mm hm CCDS SNP best RefSeq 8p11.21 zinc finger, matrin type 4
MapViewer
Recenter
Zoom in x2
Zoom in x4
Zoom in x8
Zoom out x2
Show 10M
Show 1M
Show 100K
Show 10K
Show Sequence

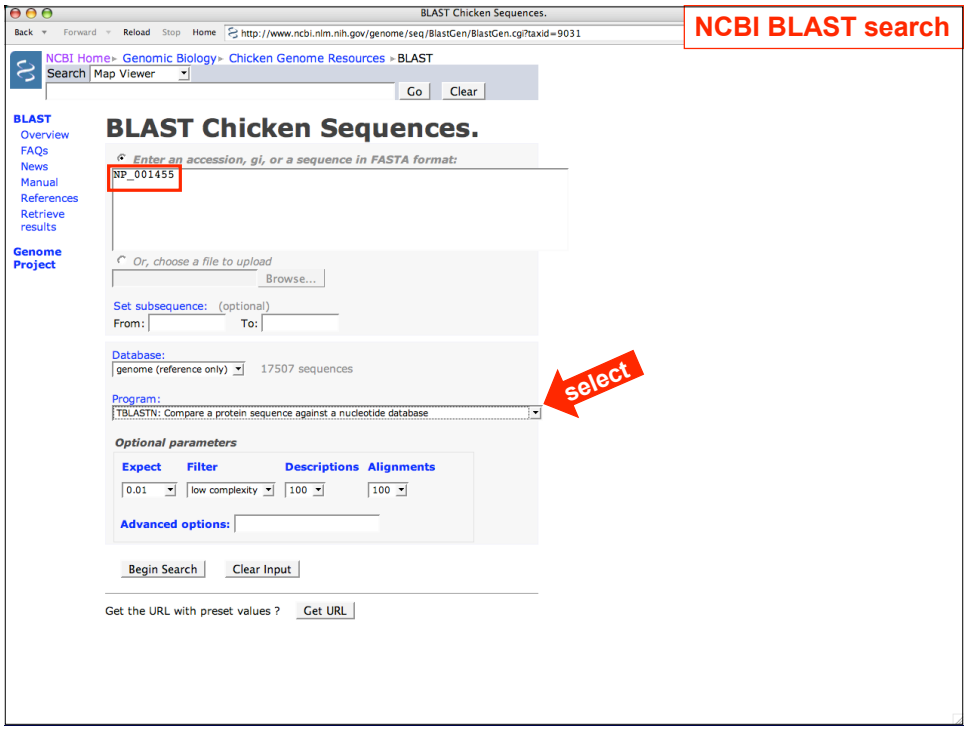

NCBI SNP map
Map Viewer
NCBI Map Viewer
Homo sapiens (human) Build 36.2 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: d8s1170 OR d8s94 - not in this view, click here to see results [clear]
Master Map: Genes On Sequence
Summary of Maps
Region Displayed: 39,609,700-39,619,700 bp
Download/View Sequence/Evidence
click
ADAM18 HGNC sv prd lev mm hm CCDS SNP best RefSeq 8p11.22 ADAM metallopeptidase domai

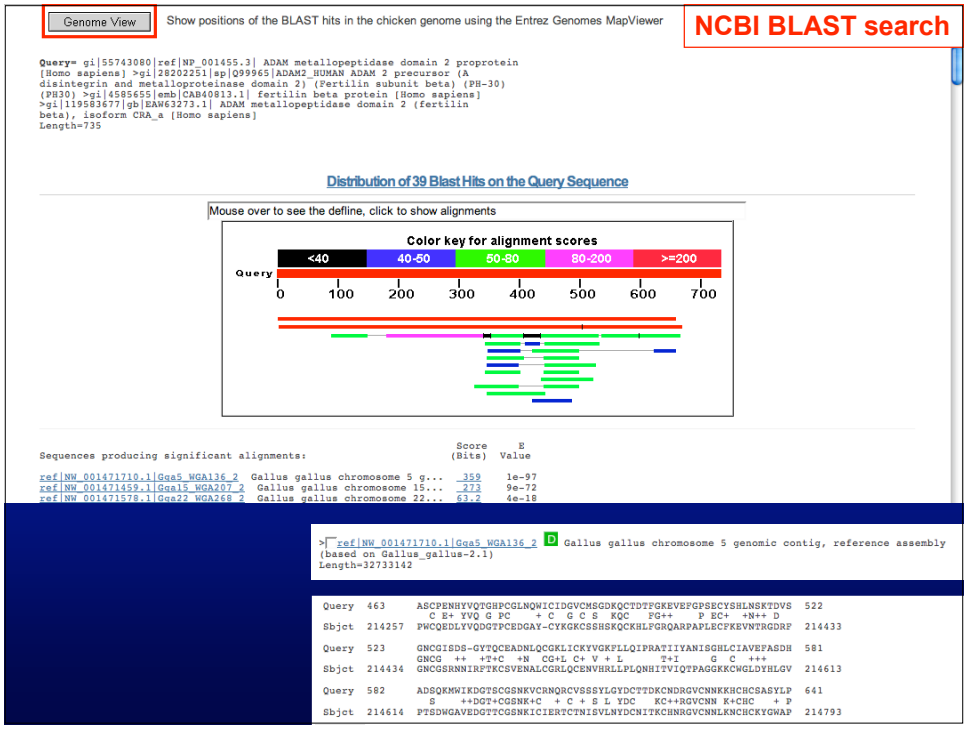

NCBI SNP map
Map Viewer
NCBI Map Viewer
Homo sapiens (human) Build 36.2 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: d8s1170 OR d8s94 - not in this view, click here to see results [clear]
Master Map: Variation
Summary of Maps
Region Displayed: 39,609,700-39,619,700 bp
Download/View Sequence/Evidence
click
12 variations
rs11785708
rs7460073
rs2167467
rs11785546
rs10108685
rs1376720
rs34064543
rs10112294
rs3824144
rs36068250
rs12708194
rs13278646
rs35854114
rs34125311
rs34185945
rs34488676
rs13248168
rs3852338
rs3852339
rs3864654
rs4532566
rs11998440
rs34787426
rs28804086

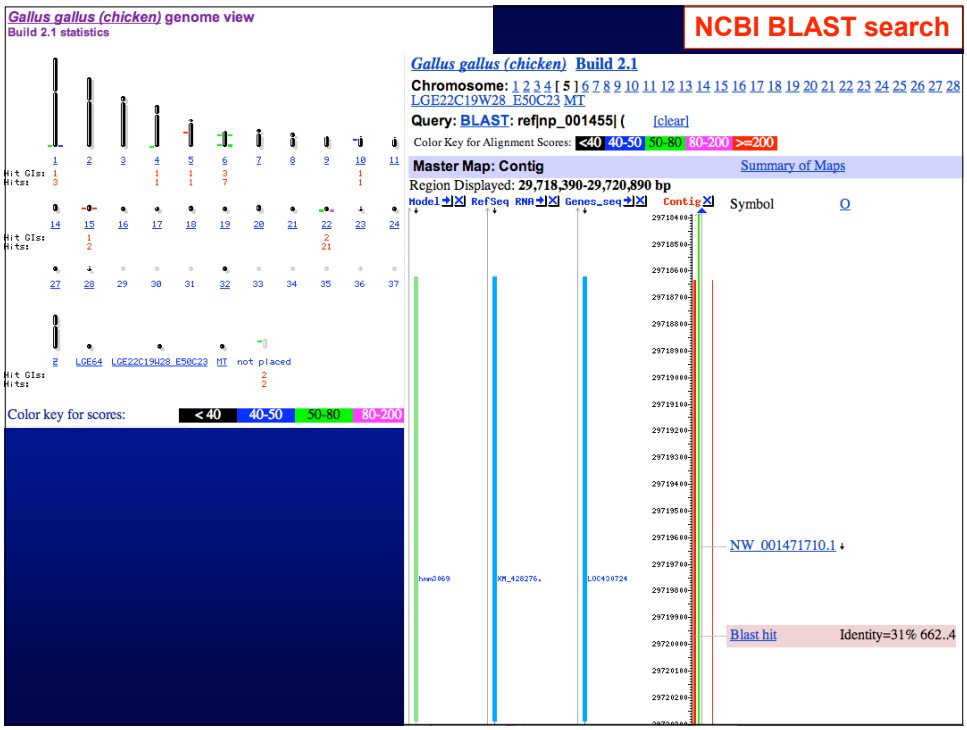

dbSNP
Single Nucleotide Polymorphism
Reference SNP(refSNP) Cluster Report: rs34064543
refSNP ID: rs34064543
Organism: human (Homo sapiens)
Molecule Type: Genomic
Created/Updated in build: 126/126
Map to Genome Build: 36.2
Allele
Variation Class: SNP: single nucleotide polymorphism
Alleles: A/T
Ancestral Allele: Not available
SNP Details are organized in the following sections:
Submission Fasta Resource GeneView Map Diversity Validation
GeneView
GeneView via analysis of contig annotation: ADAM18 ADAM metallopeptidase domain 18
reference NT_007995->NM_014237
Celera NW_923907->NM_014237
Group label | Contig->mRNA->Protein | Contig position | mRNA orientation | mRNA pos | Function | dbSNP allele | Protein residue | Codon pos | Amino acid pos
reference NT_007995->NM_014237->NP_055052 9815451 forward 667 missense T Ser [S] 1 223
contig reference A Thr [T] 1 223
Celera NW_923907->NM_014237->NP_055052 27014721 forward 667 missense T Ser [S] 1 223
contig reference A Thr [T] 1 223

NCBI

Find a chicken homolog of a human protein

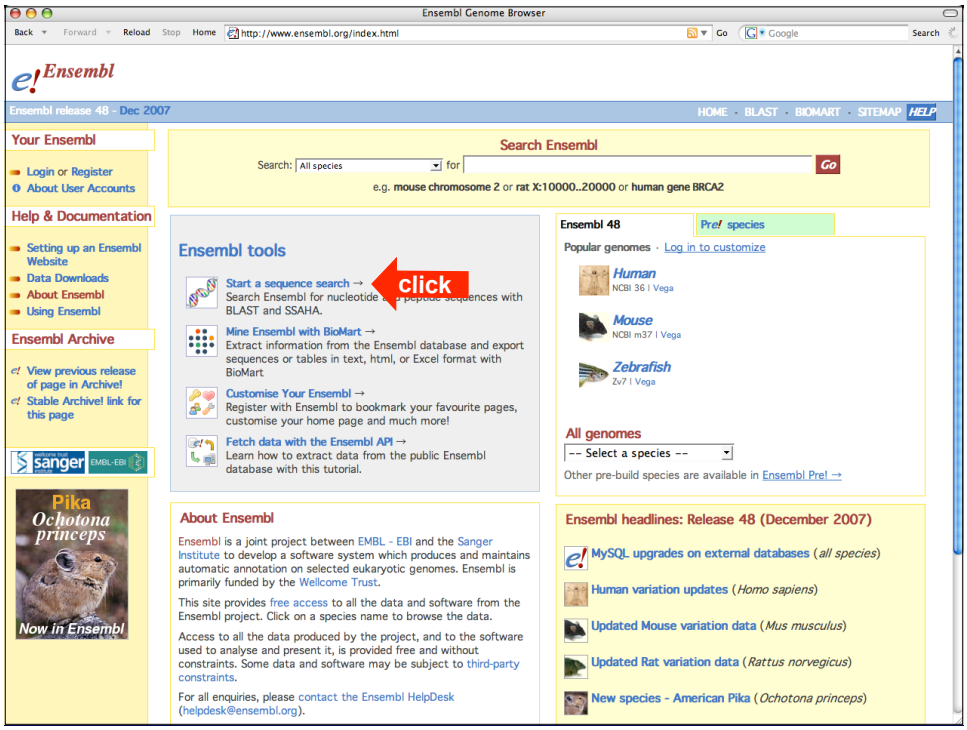


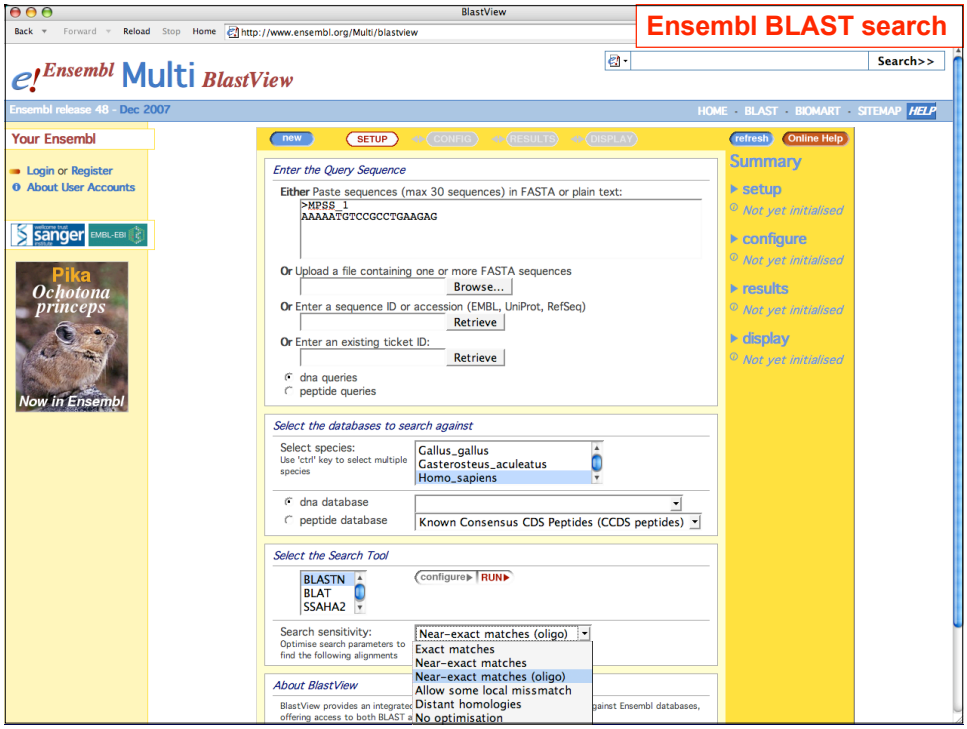

Genome View
Show positions of the BLAST hits in the chicken genome using the Entrez Genomes MapViewer
NCBI BLAST search
Distribution of 39 Blast Hits on the Query Sequence
Mouse over to see the defline, click to show alignments
Color key for alignment scores
<40
40-50
50-80
80-200
>=200
Query
0
100
200
300
400
500
600
700
Sequences producing significant alignments:
Score (Bits)
E Value
ref|NW_001471710.1|Gga5_WGA136_2 Gallus gallus chromosome 5 g... 359 1e-97
ref|NW_001471459.1|Gga15_WGA207_2 Gallus gallus chromosome 15... 273 9e-72
ref|NW_001471578.1|Gga22_WGA268_2 Gallus gallus chromosome 22... 63.2 4e-18
ref|NW_001471710.1|Gga5_WGA136_2 Gallus gallus chromosome 5 genomic contig, reference assembly
(based on Gallus_gallus-2.1)
Length=32733142

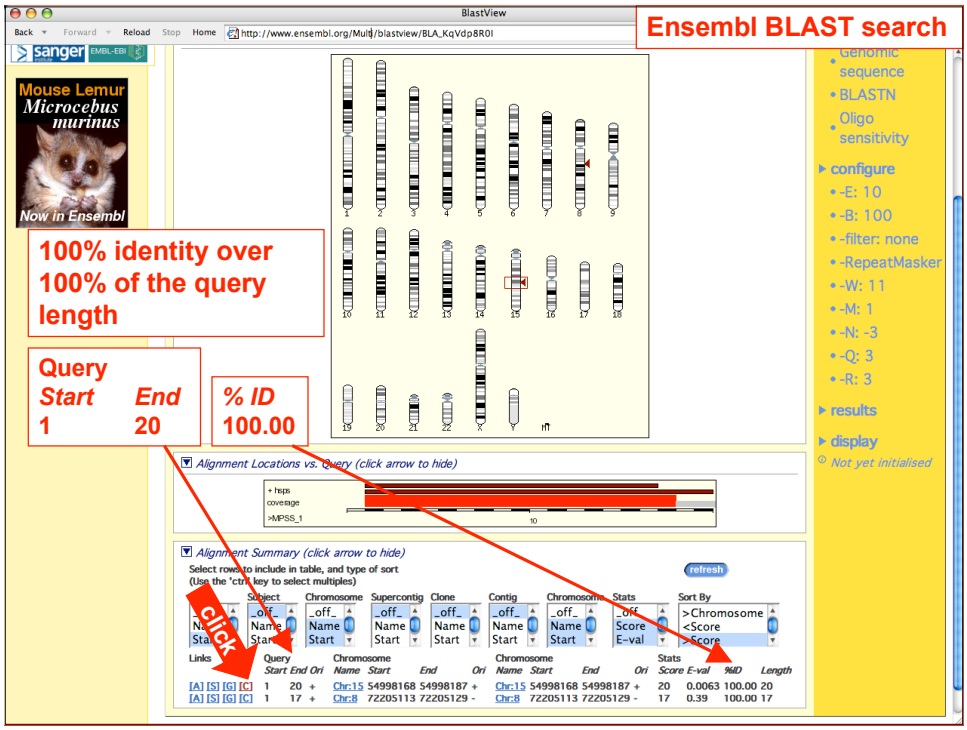

Gallus gallus (chicken) genome view
Build 2.1 statistics
NCBI BLAST search
Gallus gallus (chicken) Build 2.1
Chromosome: 1 2 3 4 [5] 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 LGE22C19W28_E50C23 MT
Query: BLAST: ref|np_001455| ([clear]
Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200
Master Map: Contig
Summary of Maps
Region Displayed: 29,718,390-29,720,890 bp
Model RefSeq RNA Genes_seq Contig
Symbol
Hit GIs:
Hits:
Color key for scores:
< 40
40-50
50-80
80-200
NW_001471710.1
Blast hit
Identity=31% 662..4

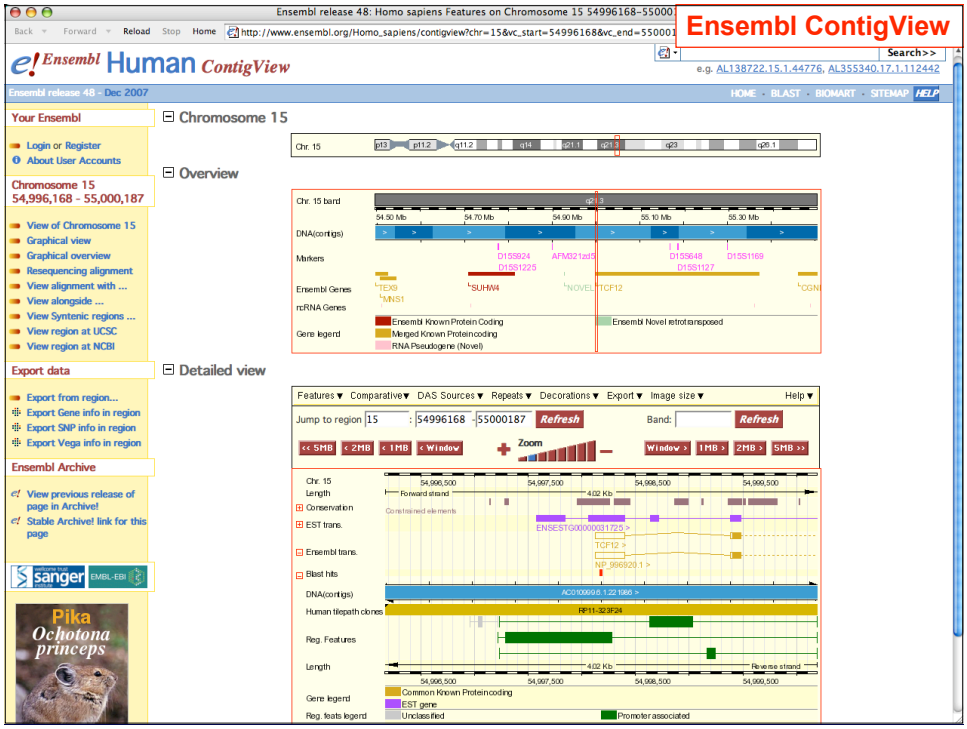
Ensembl

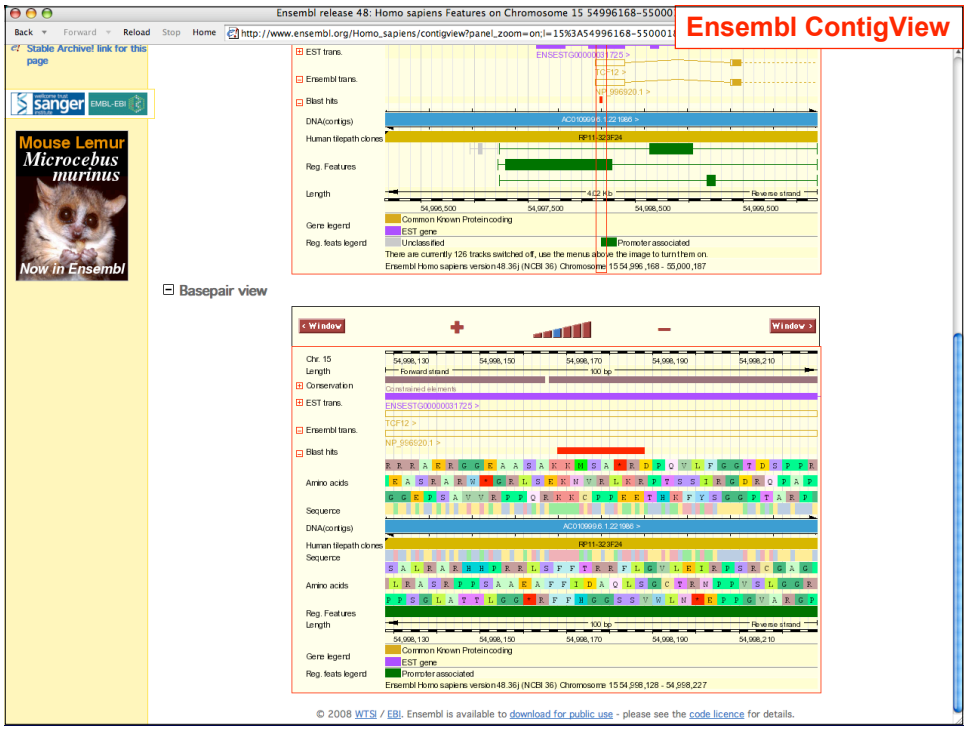
Identify genes that overlap with an oligo tag




Ensembl BLAST search
BlastView
http://www.ensembl.org/Multi/blastview
Ensembl Multi BlastView
Search>>
Ensembl release 48 - Dec 2007
HOME - BLAST - BIOMART - SITEMAP HELP
Your Ensembl
Login or Register
About User Accounts
Pika
Ochotona princeps
Now in Ensembl
new
SETUP
CONFIG
RESULTS
DISPLAY
refresh
Online Help
Summary
setup
Not yet initialised
configure
Not yet initialised
results
Not yet initialised
display
Not yet initialised
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
Browse...
Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)
Retrieve
Or Enter an existing ticket ID:
Retrieve
dna queries
peptide queries
Select the databases to search against
Select species:
Use 'ctrl' key to select multiple species
Gallus_gallus
Gasterosteus_aculeatus
Homo_sapiens
dna database
peptide database
Known Consensus CDS Peptides (CCDS peptides)
Select the Search Tool
BLASTN
BLAT
SSAHA2
configure
RUN
Search sensitivity:
Optimise search parameters to find the following alignments
Near-exact matches (oligo)
Exact matches
Near-exact matches
Near-exact matches (oligo)
Allow some local missmatch
Distant homologies
No optimisation
About BlastView

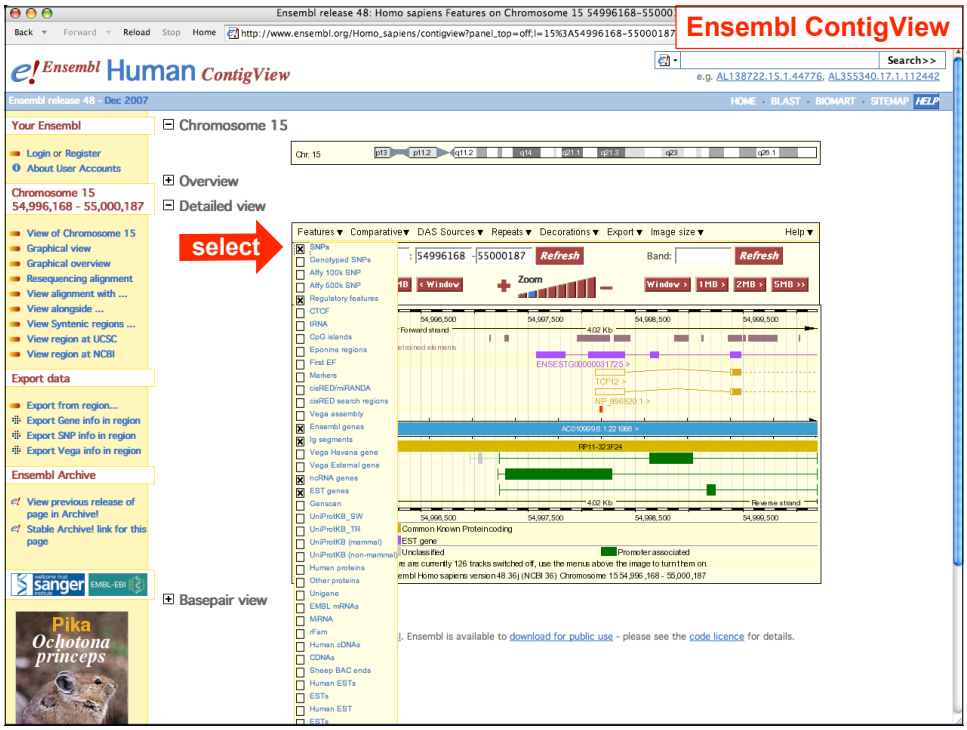

Ensembl BLAST search
BlastView
http://www.ensembl.org/Multi/blastview/BLA_KqVdp8R0I
Mouse Lemur
Microcebus murinus
Now in Ensembl
100% identity over 100% of the query length
Query
Start
End
1
20
% ID
100.00
click
Genomic sequence
BLASTN
Oligo sensitivity
configure
-E: 10
-B: 100
-filter: none
-RepeatMasker
-W: 11
-M: 1
-N: -3
-Q: 3
-R: 3
results
display
Not yet initialised
Alignment Locations vs. Query (click arrow to hide)
Alignment Summary (click arrow to hide)
Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)
refresh
Sort By
>Chromosome
<Score
>Score
Links
Query
Chromosome
Stats
[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

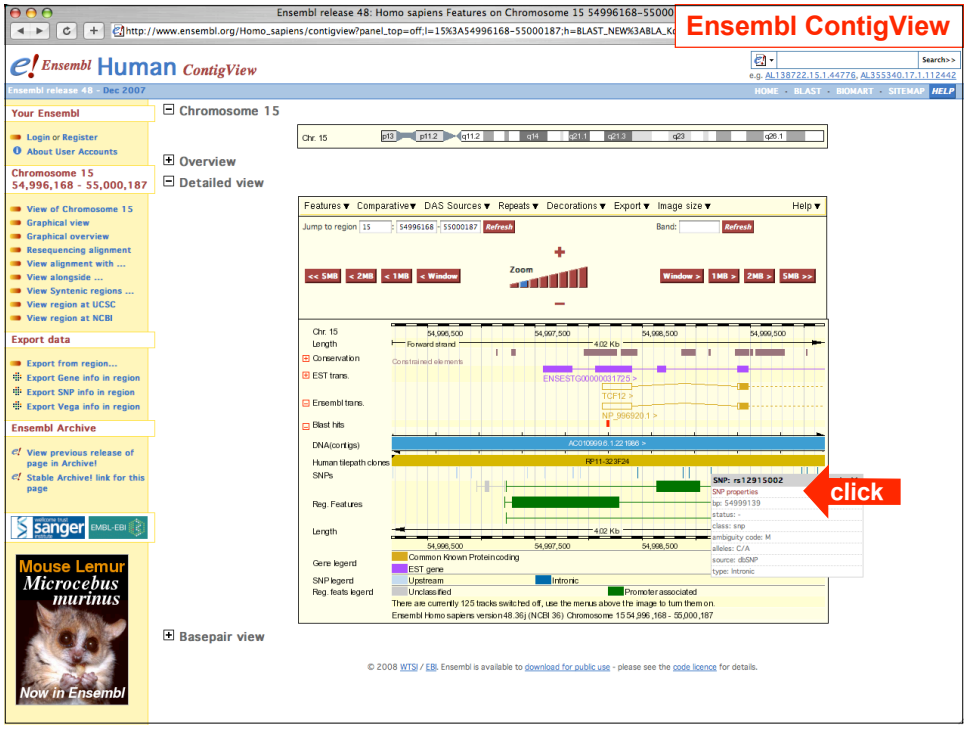

Ensembl ContigView
Ensembl release 48: Homo sapiens Features on Chromosome 15 54996168-55000
Ensembl Human ContigView
Chromosome 15
Overview
Detailed view
Your Ensembl
Export data
Ensembl Archive
Pika
Ochotona princeps

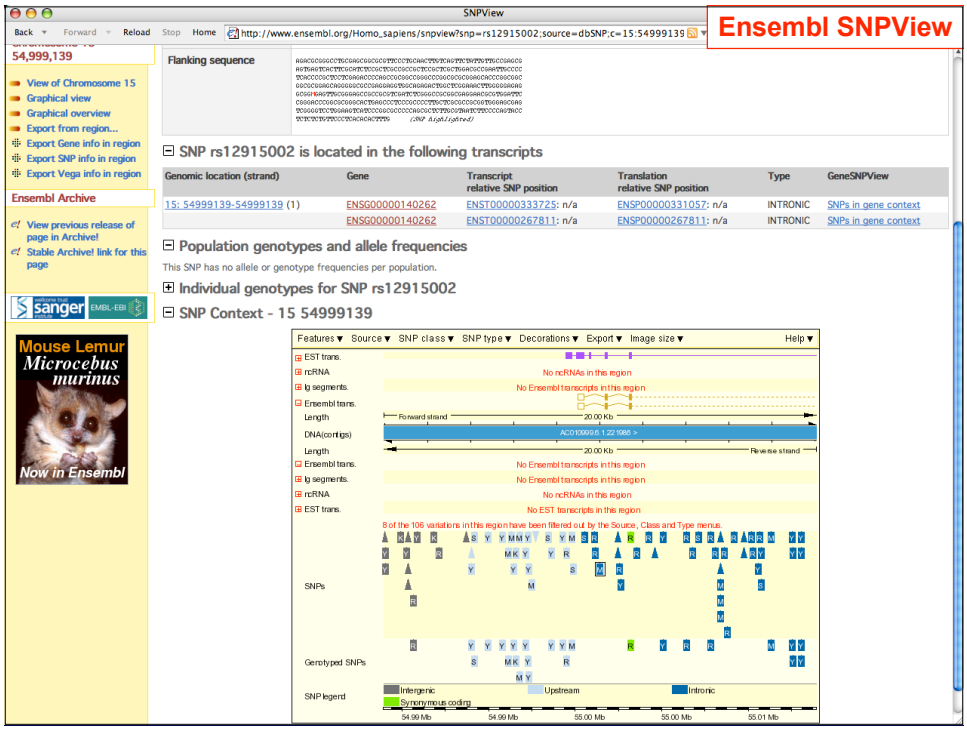

Ensembl ContigView
Ensembl release 48: Homo sapiens Features on Chromosome 15 54996168-55000
Mouse Lemur
Microcebus murinus
Now in Ensembl
Basepair view

Ensembl

Add features to the ContigView

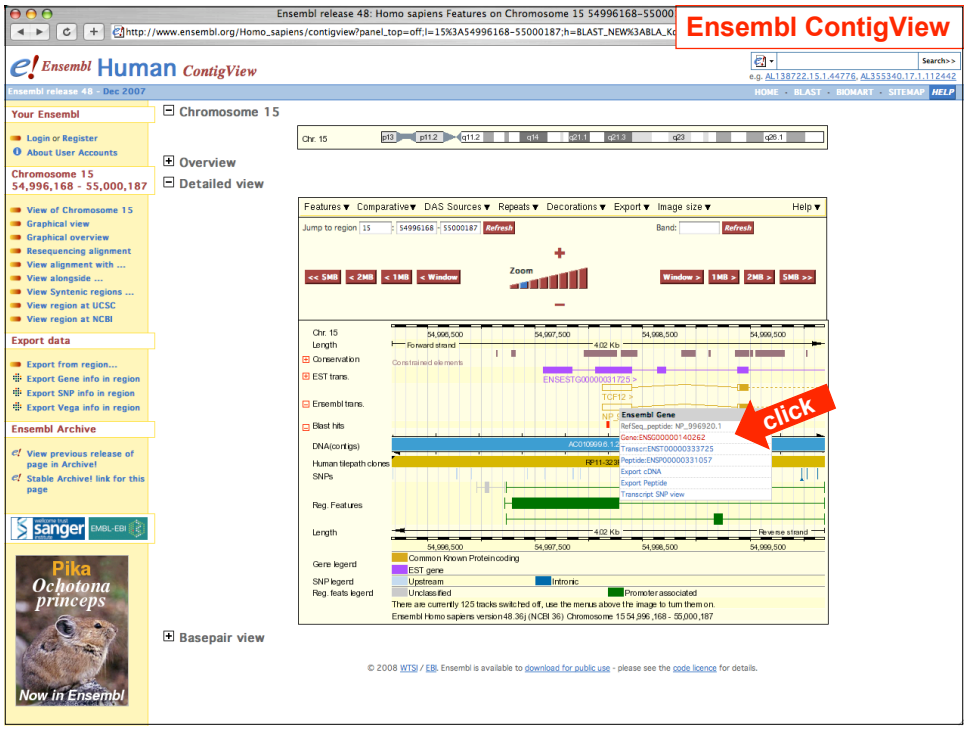


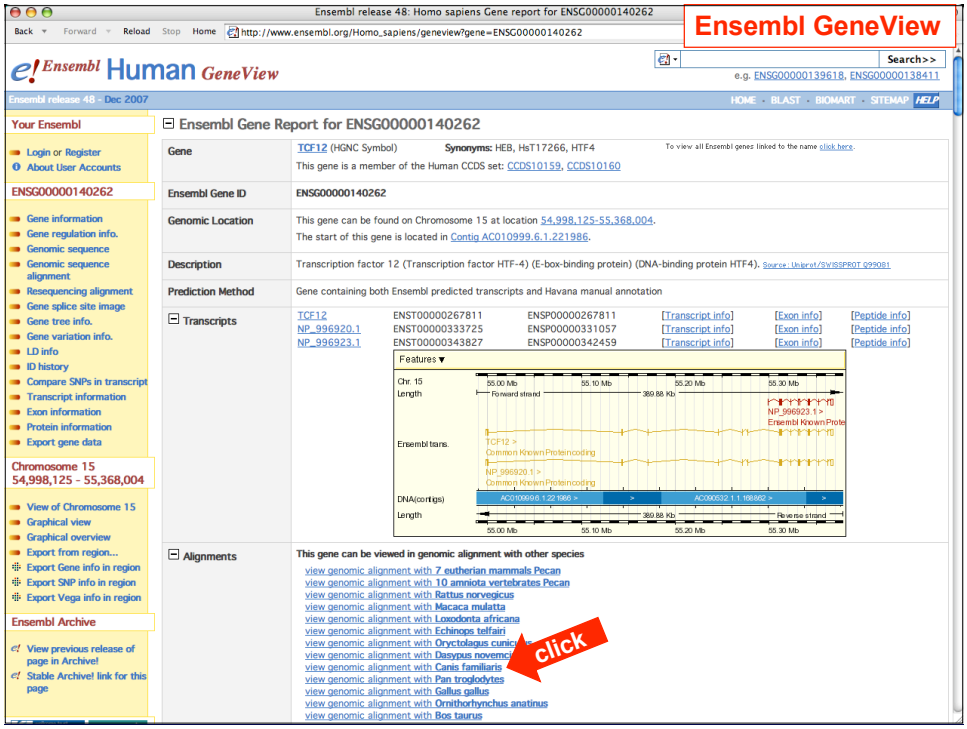

Ensembl ContigView
Ensembl release 48: Homo sapiens Features on Chromosome 15 54996168-55000
Ensembl Human ContigView
Chromosome 15
Overview
Detailed view
Basepair view
click
Mouse Lemur
Microcebus murinus
Now in Ensembl

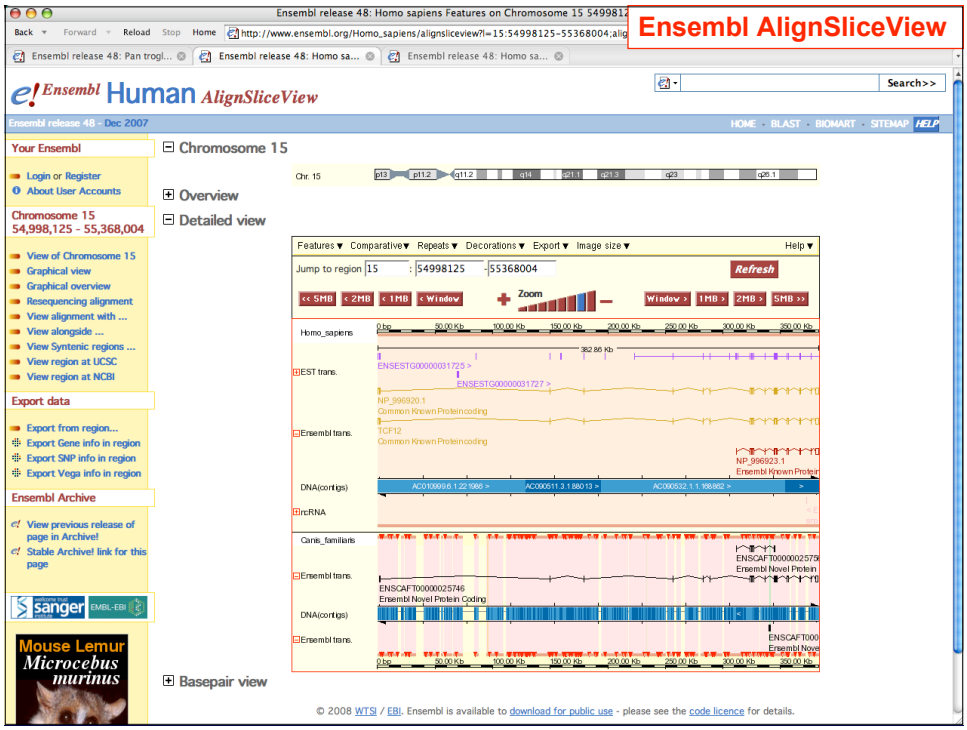

Ensembl SNPView
SNPView
Flanking sequence
SNP rs12915002 is located in the following transcripts
Population genotypes and allele frequencies
This SNP has no allele or genotype frequencies per population.
Individual genotypes for SNP rs12915002
SNP Context - 15 54999139
Mouse Lemur
Microcebus murinus
Now in Ensembl

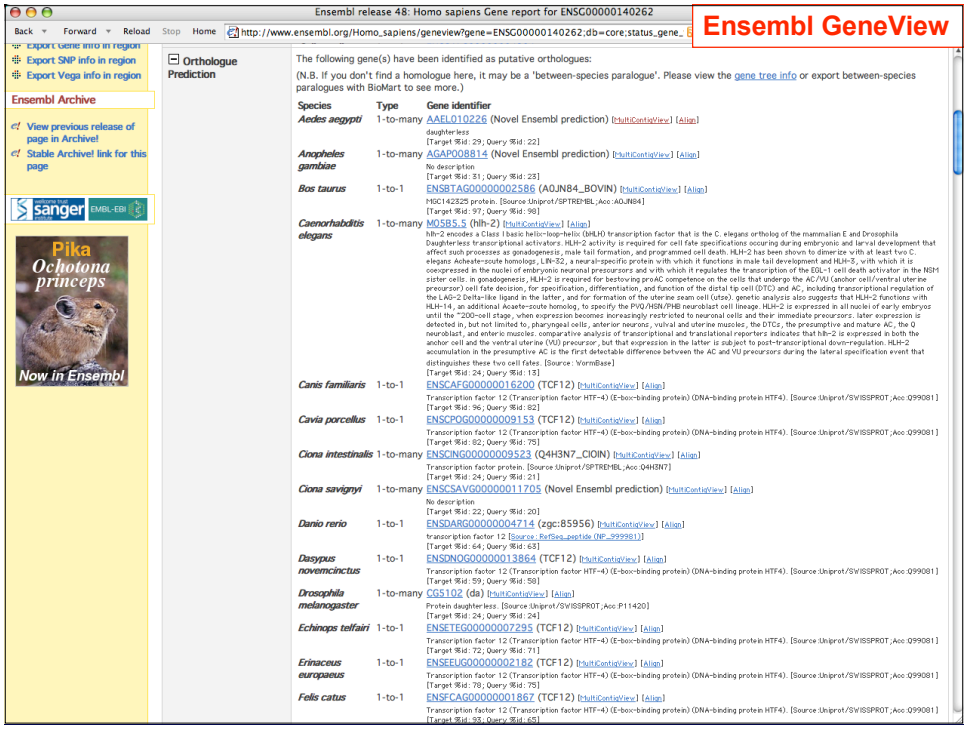
Ensembl

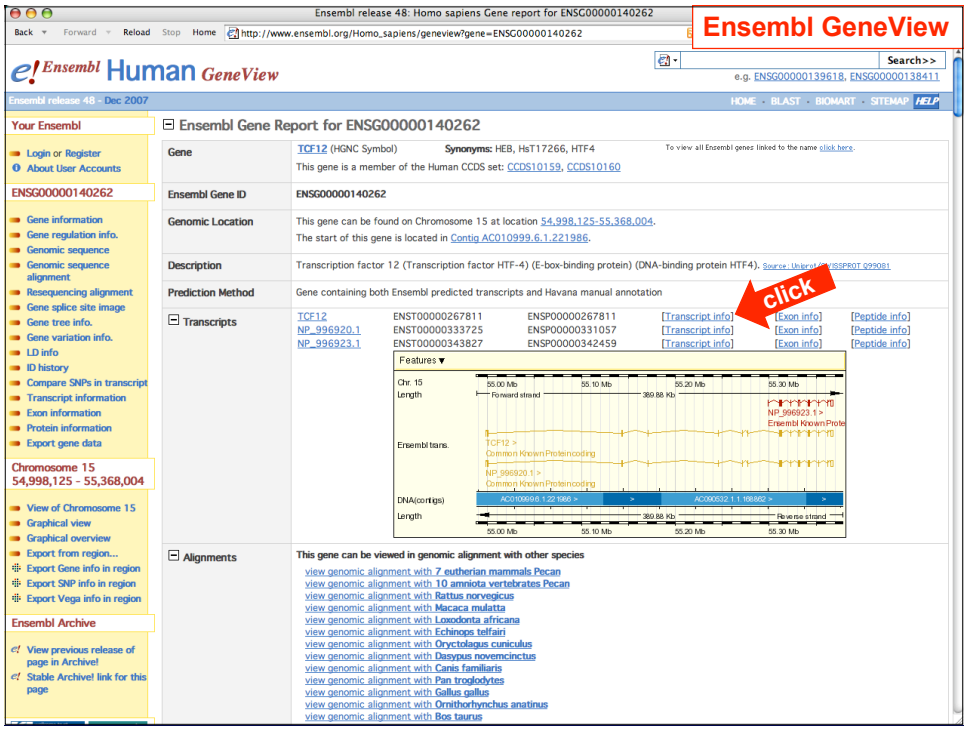
Get information about the gene, transcripts, exons, and protein

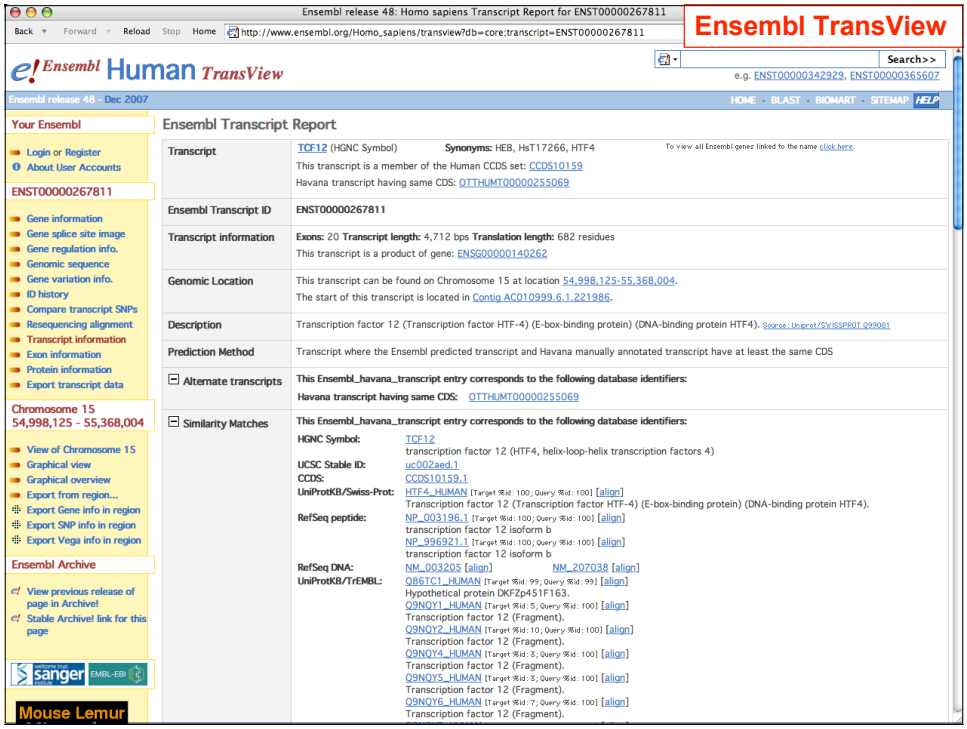


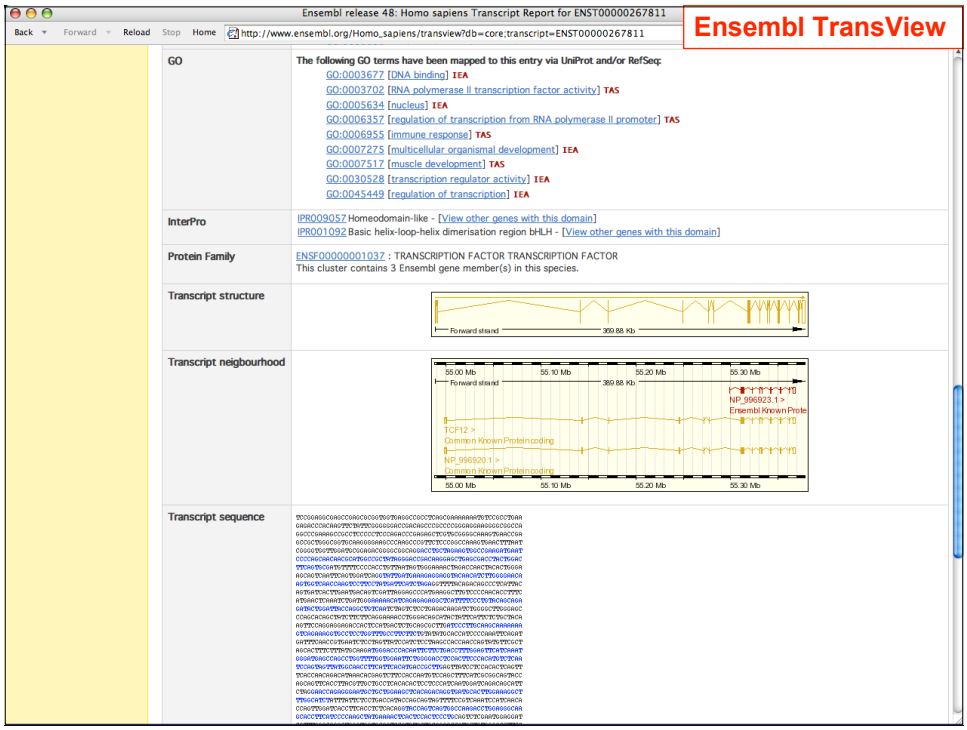

Ensembl GeneView
Ensembl release 48: Homo sapiens Gene report for ENSG00000140262
Ensembl Human GeneView
Ensembl Gene Report for ENSG00000140262
click

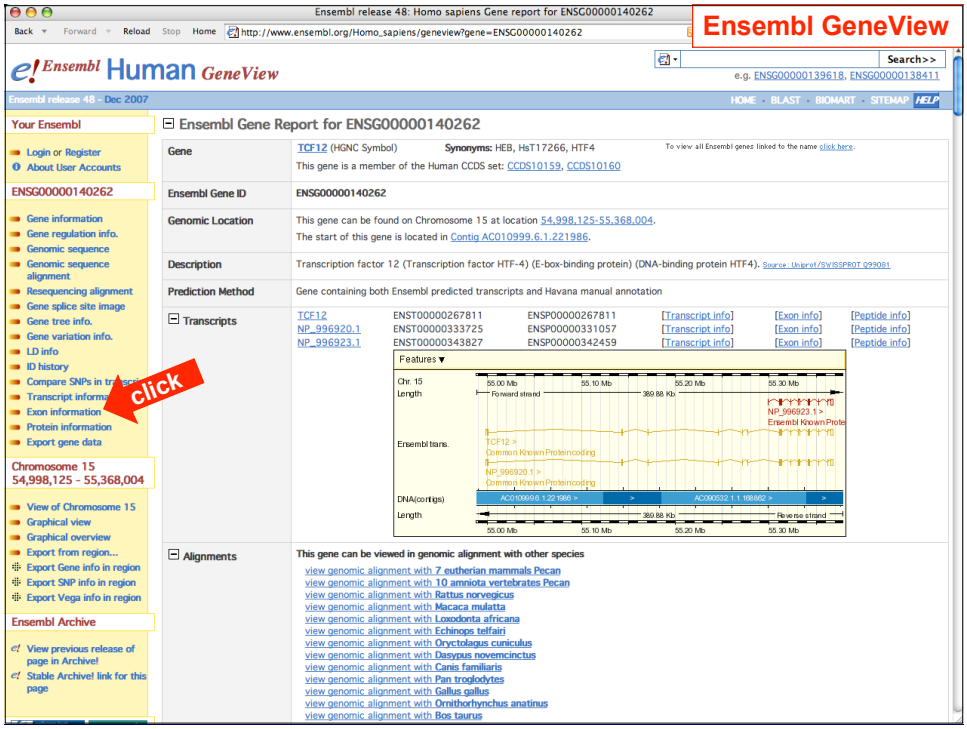

Ensembl AlignSliceView
Ensembl release 48: Homo sapiens Features on Chromosome 15
Ensembl Human AlignSliceView
Chromosome 15
Overview
Detailed view
Basepair view

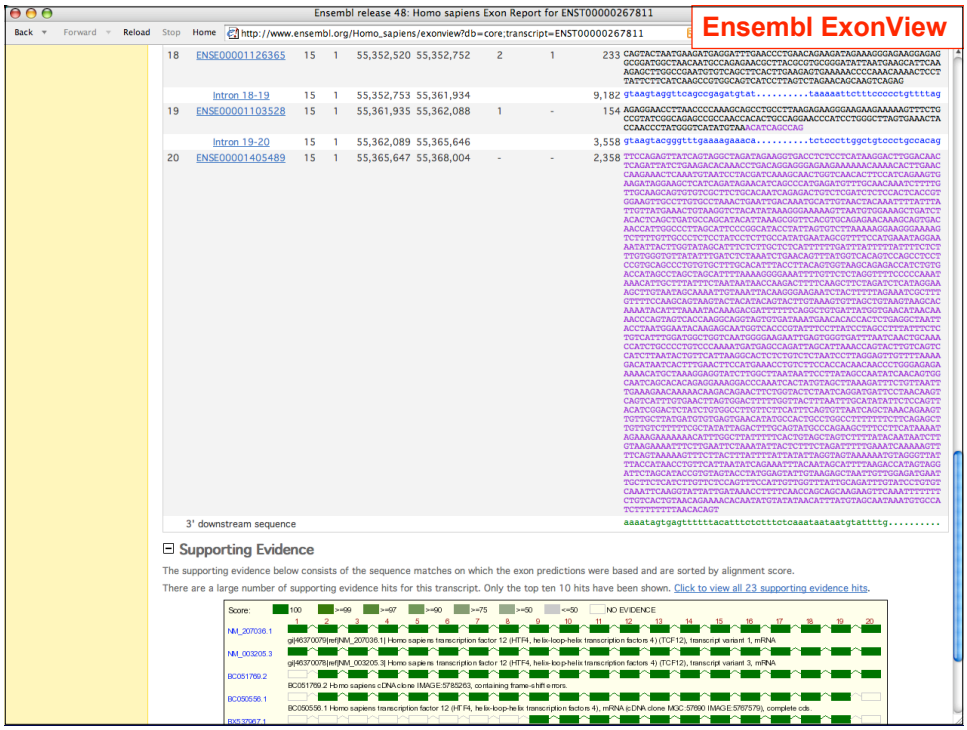

Ensembl release 48: Homo sapiens Gene report for ENSG00000140262
Ensembl GeneView
Orthologue Prediction
The following gene(s) have been identified as putative orthologues:
Export SNP info in region
Export Vega info in region
Ensembl Archive
View previous release of page in Archive!
Stable Archive! link for this page
Now in Ensembl

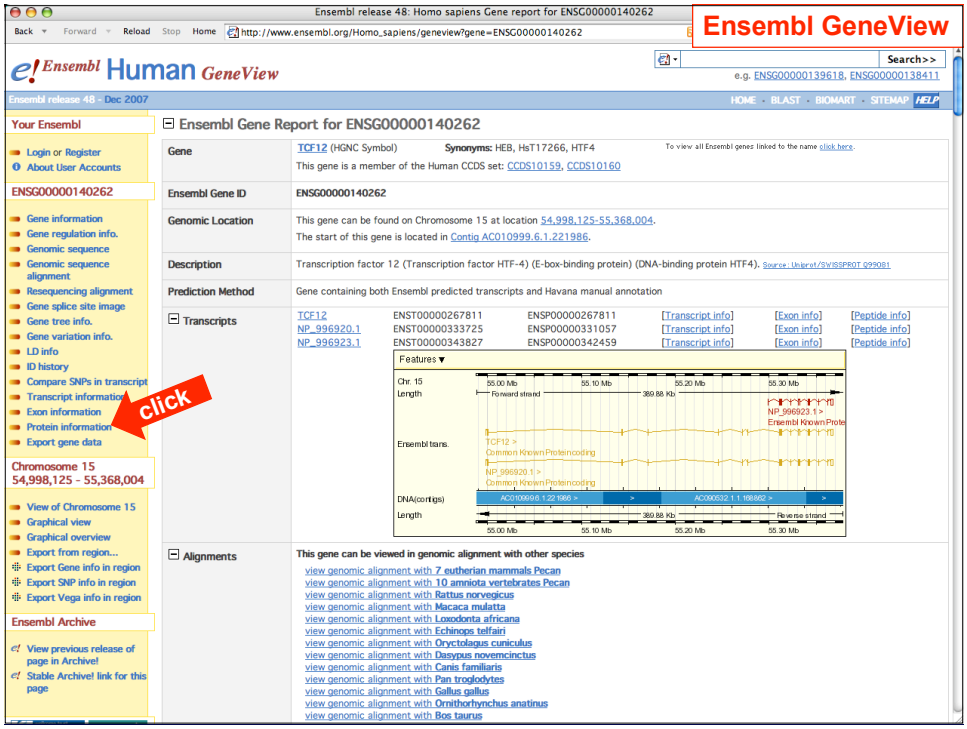

Ensembl release 48: Homo sapiens Gene report for ENSG00000140262
Ensembl GeneView
Ensembl Human GeneView
Ensembl Gene Report for ENSG00000140262
Gene
Ensembl Gene ID
Genomic Location
Description
Prediction Method
Transcripts
click
Alignments
Your Ensembl
Chromosome 15
54,998,125 - 55,368,004

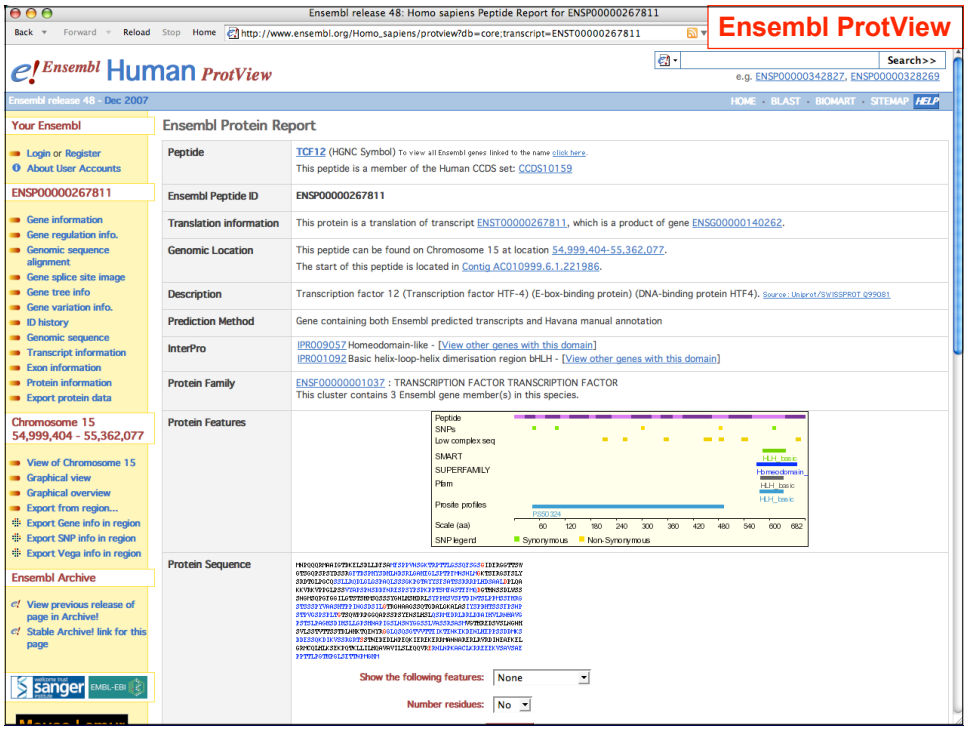

Ensembl release 48: Homo sapiens Transcript Report for ENST00000267811
Ensembl TransView
http://www.ensembl.org/Homo_sapiens/transview?db=core;transcript=ENST00000267811
Ensembl Human TransView
Ensembl Transcript Report
Transcript
Ensembl Transcript ID
ENST00000267811
Transcript information
Genomic Location
Description
Prediction Method
Alternate transcripts
Similarity Matches
Your Ensembl
Chromosome 15
54,998,125 - 55,368,004
Ensembl Archive
Mouse Lemur


Ensembl release 48: Homo sapiens Transcript Report for ENST00000267811
Ensembl TransView
GO
InterPro
Protein Family
Transcript structure
Transcript neigbourhood
Transcript sequence


Ensembl GeneView
Ensembl release 48: Homo sapiens Gene report for ENSG00000140262
http://www.ensembl.org/Homo_sapiens/geneview?gene=ENSG00000140262
Ensembl Human GeneView
Ensembl Gene Report for ENSG00000140262
click

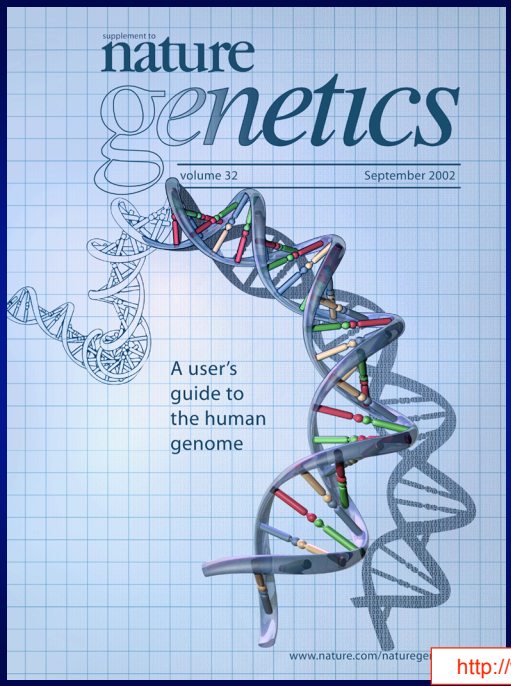

Ensembl ExonView
Ensembl release 48: Homo sapiens Exon Report for ENST00000267811
http://www.ensembl.org/Homo_sapiens/exonview?db=core;transcript=ENST00000267811
Supporting Evidence


Ensembl GeneView
Ensembl release 48: Homo sapiens Gene report for ENSG00000140262
http://www.ensembl.org/Homo_sapiens/geneview?gene=ENSG00000140262
Ensembl Human GeneView
Ensembl release 48 - Dec 2007
Ensembl Gene Report for ENSG00000140262
Gene
Ensembl Gene ID
Genomic Location
Description
Prediction Method
Transcripts
Alignments
click
Your Ensembl
ENSG00000140262
Chromosome 15
54,998,125 - 55,368,004
Ensembl Archive


Ensembl ProtView
Ensembl release 48: Homo sapiens Peptide Report for ENSP00000267811
http://www.ensembl.org/Homo_sapiens/protview?db=core;transcript=ENST00000267811
Ensembl Human ProtView
Ensembl release 48 - Dec 2007
Ensembl Protein Report
Peptide
Ensembl Peptide ID
Translation information
Genomic Location
Description
Prediction Method
InterPro
Protein Family
Protein Features
Protein Sequence
Your Ensembl
ENSP00000267811
Chromosome 15
54,999,404 - 55,362,077
Ensembl Archive
Show the following features:
Number residues:

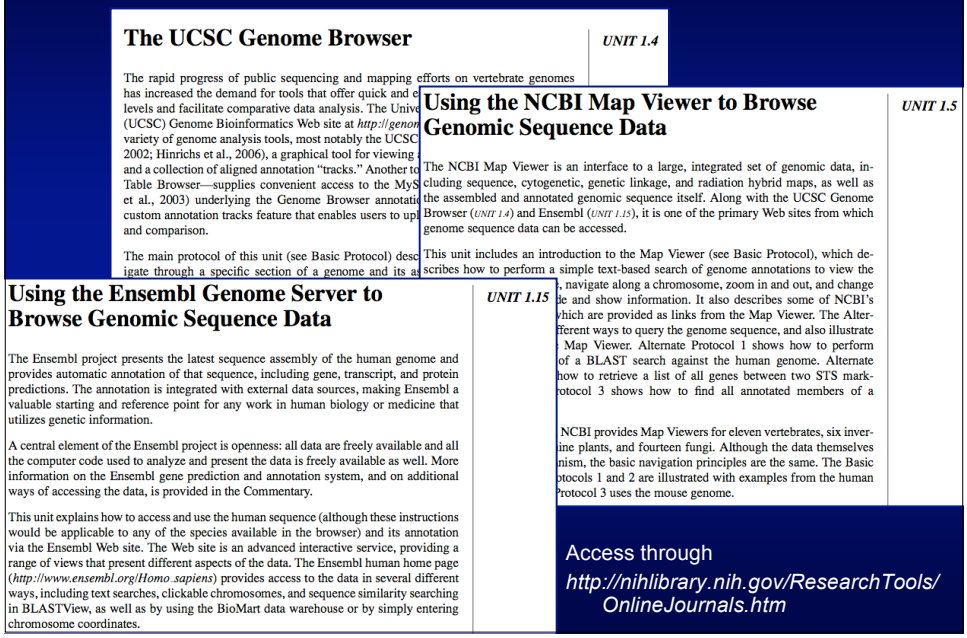
Additional resources

- 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/focsup_index.html

Current Protocols in Bioinformatics



The UCSC Genome Browser *UNIT 1.4*

The rapid progress of public sequencing and mapping efforts on vertebrate genomes has increased the demand for tools that offer quick and e levels and facilitate comparative data analysis. The Unive (UCSC) Genome Bioinformatics Web site at *http://genom* variety of genome analysis tools, most notably the UCSC 2002; Hinrichs et al., 2006), a graphical tool for viewing and a collection of aligned annotation "tracks." Another to Table Browser—supplies convenient access to the MyS et al., 2003) underlying the Genome Browser annotati custom annotation tracks feature that enables users to upl and comparison.

The main protocol of this unit (see Basic Protocol) desc igate through a specific section of a genome and its as

Using the NCBI Map Viewer to Browse Genomic Sequence Data *UNIT 1.5*

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.

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 , navigate along a chromosome, zoom in and out, and change de and show information. It also describes some of NCBI's which are provided as links from the Map Viewer. The Alter- fferent ways to query the genome sequence, and also illustrate Map Viewer. Alternate Protocol 1 shows how to perform of a BLAST search against the human genome. Alternate how to retrieve a list of all genes between two STS mark- rotocol 3 shows how to find all annotated members of a

NCBI provides Map Viewers for eleven vertebrates, six inver- line plants, and fourteen fungi. Although the data themselves nism, the basic navigation principles are the same. The Basic otocols 1 and 2 are illustrated with examples from the human rotocol 3 uses the mouse genome.

Using the Ensembl Genome Server to Browse Genomic Sequence Data *UNIT 1.15*

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.

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.

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.

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

Current Topics in Genome Analysis

Next Lecture:

Evolutionary Analysis

Fiona Brinkman, Ph.D.
Simon Fraser University