NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research

Current Topics in Genome Analysis
Spring 2010

Week 4: Mining Genomic Sequence Data

Tyra Wolfsberg, Ph.D.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES | NATIONAL INSTITUTES OF HEALTH | genome.gov/DIR

Accessing the public genome sequence data

UCSC's Genome Browser ("Golden Path")
http://genome.ucsc.edu

Ensembl
http://www.ensembl.org

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

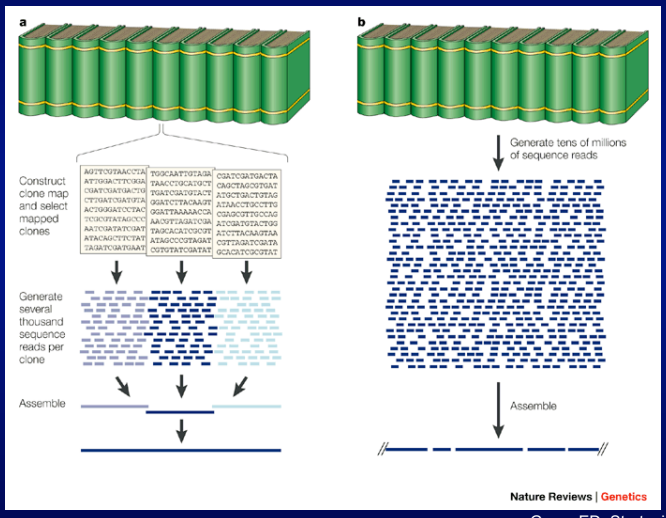
Types of data integrated in genome browsers

- 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 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	Feb 2009/hg19/ GRCh37	Build 37.1	GRCh37
Mouse	Yes	July 2007/mm9/Build 37	Build 37.1	Build 37
Dog	Yes	May 2005 /canFam 2.0	Build 2.1/ CanFam 2.0	CanFam 2.0
Zebrafish	NO	Dec 2008/danRer6/ Zv8	Zv7/build 3.1	Zv8
Rhesus	Yes	Jan 2006/rheMac2/ v. 1.0, Mmul_051212	Build 1.1/ v.1.0, Mmul_051 212	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

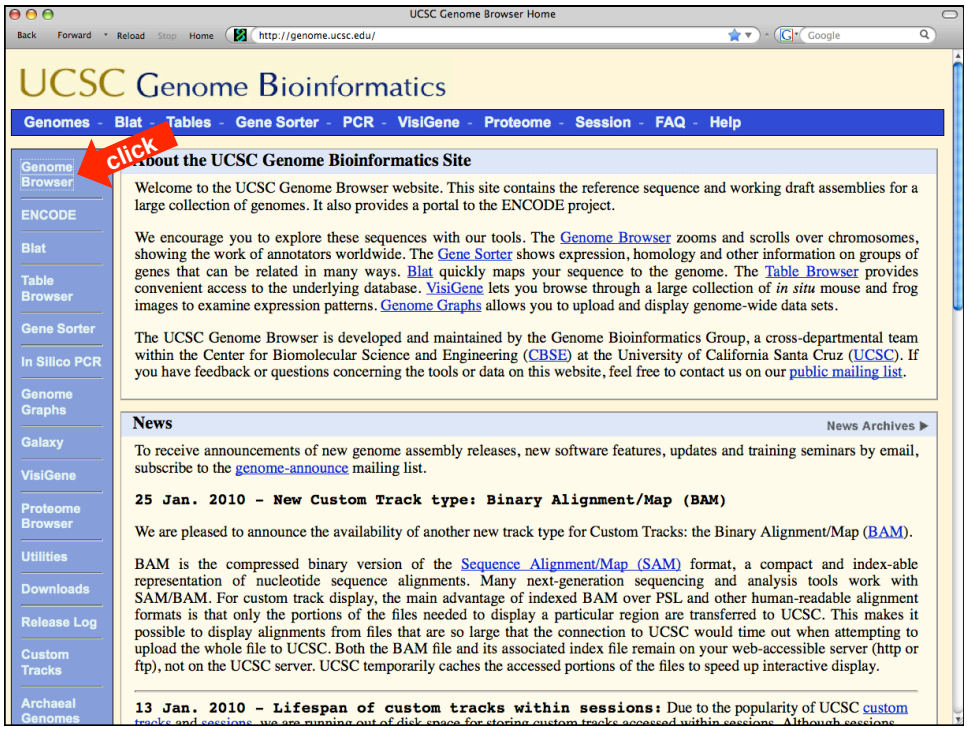
http://www.ncbi.nlm.nih.gov/RefSeq/key.html

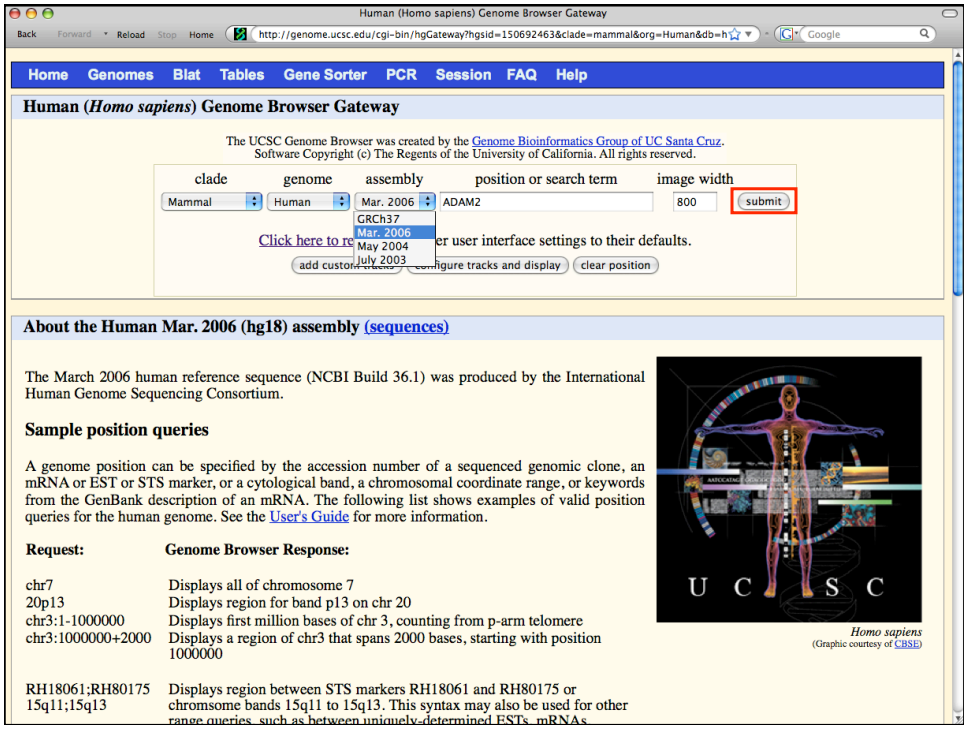
Beta actin mRNA RefSeq

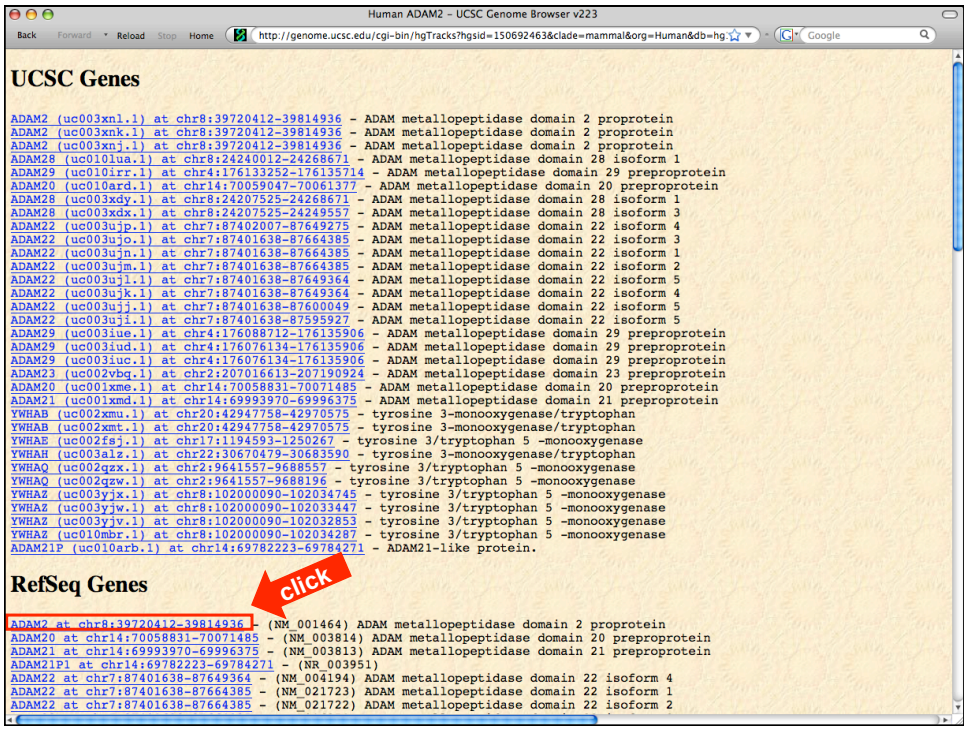
```
LOCUS       NM_001101               1852 bp    mRNA    linear   PRI 27-DEC-2009
DEFINITION  Homo sapiens actin, beta (ACTB), mRNA.
ACCESSION   NM_001101
VERSION     NM_001101.3  GI:168480144
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1852)
  AUTHORS   Yamaguchi,H., Shiraishi,M., Fukami,K., Tanabe,A., Ikeda-Matsuo,Y.,
            Naito,Y. and Sasaki,Y.
  TITLE     MARCKS regulates lamellipodia formation induced by IGF-I via
            association with PIP2 and beta-actin at membrane microdomains
  JOURNAL   J. Cell. Physiol. 220 (3), 748-755 (2009)
   PUBMED   19475567

COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from AK130157.1 and BC009636.1.
            On Feb 22, 2008 this sequence version replaced gi:5016088.

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

            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             85..1212
                     /gene="ACTB"
                     /gene_synonym="PS1TP5BP1"
                     /note="beta cytoskeletal actin; PS1TP5-binding protein 1;
                     actin, cytoplasmic 1"
                     /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 accgccgaga ccgcgtccgc cccgcgagca cagagcctcg cctttgccga tccgccgccc
       61 gtccacaccc gccgccagct caccatggat gatgatatcg ccgcgctcgt cgtcgacaac
      121 ggctccggca tgtgcaaggc cggcttcgcg ggcgacgatg ccccccgggc cgtcttcccc
```

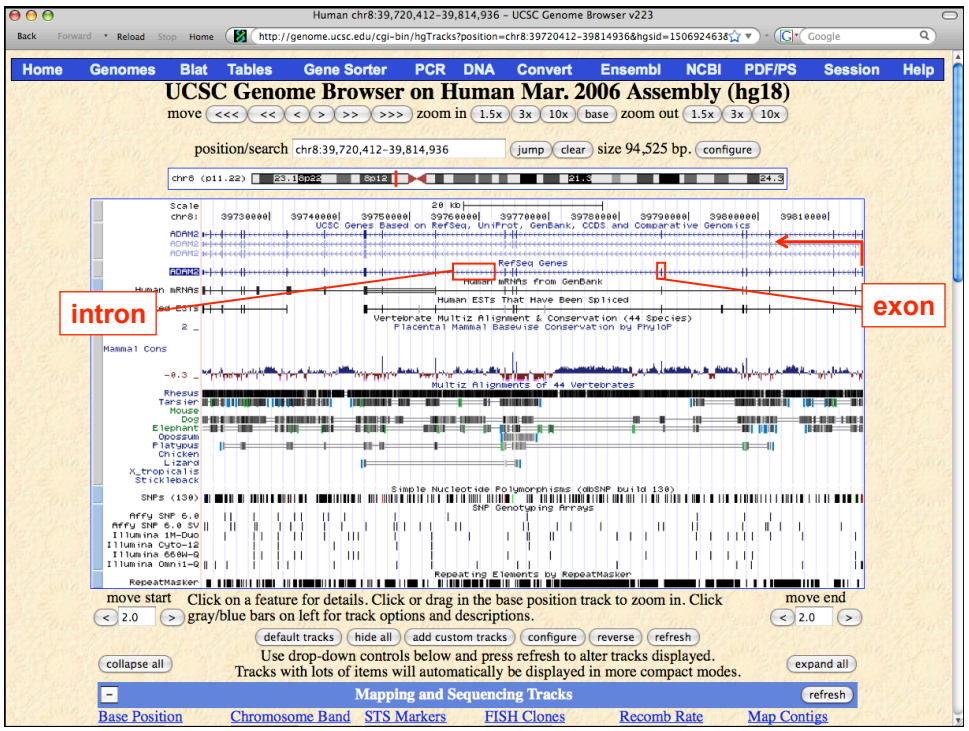
UCSC

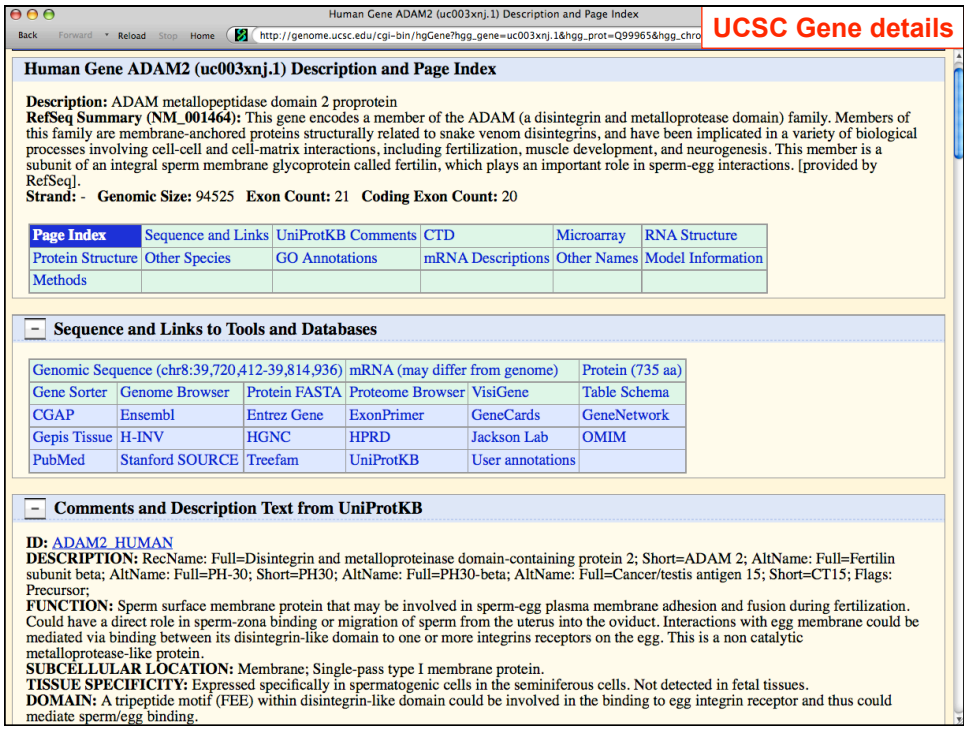
View a region in the genome by querying with a gene symbol

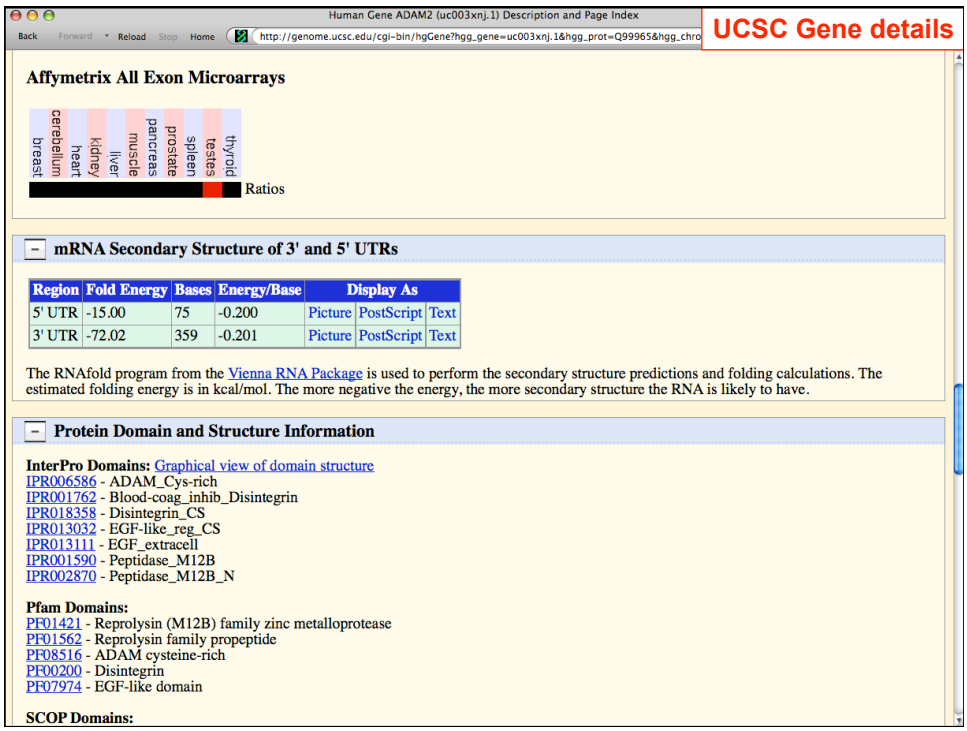


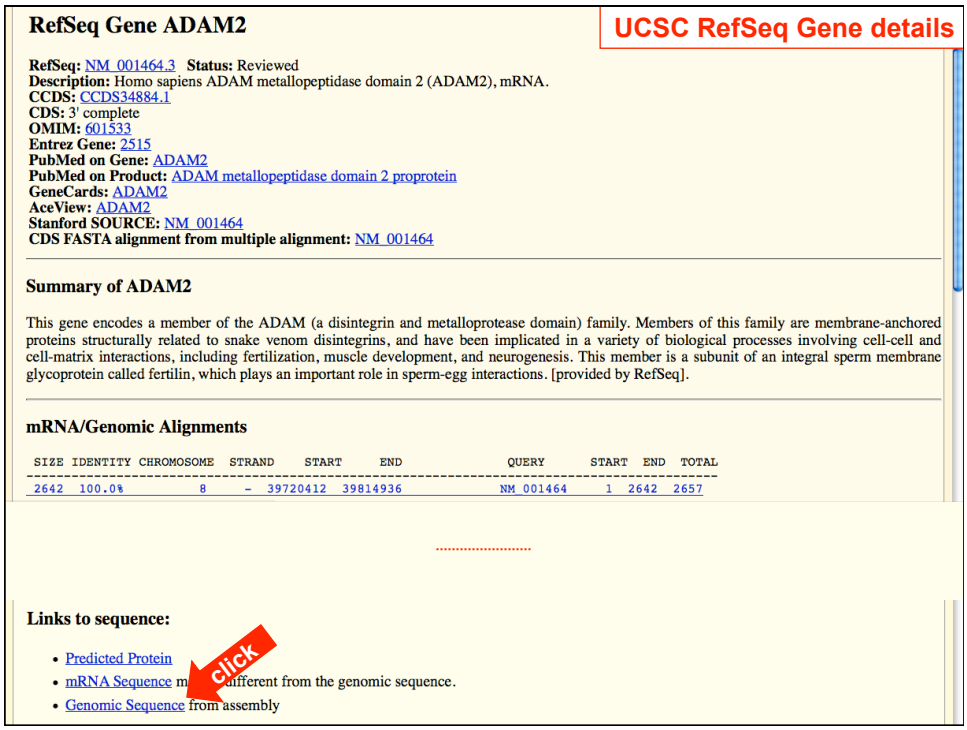

Human (Homo sapiens) Genome Browser Gateway
Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help
Human (Homo sapiens) Genome Browser Gateway
clade genome assembly position or search term image width
Mammal Human Mar. 2006 ADAM2 800 submit
GRCh37
Mar. 2006
May 2004
July 2003
clear position
About the Human Mar. 2006 (hg18) assembly (sequences)
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
RH18061;RH80175 15q11;15q13 Displays region between STS markers RH18061 and RH80175 or chromsome bands 15q11 to 15q13. This syntax may also be used for other
U C S C
Homo sapiens

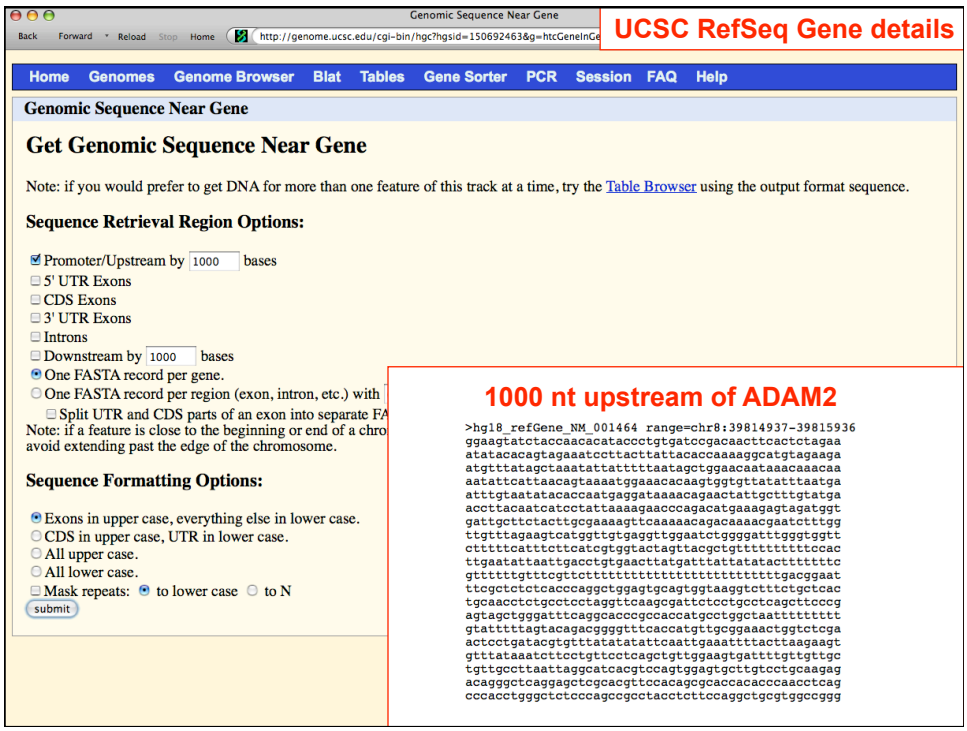

Human ADAM2 - UCSC Genome Browser v223
UCSC Genes
ADAM2 (uc003xnl.1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein
ADAM2 (uc003xnk.1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein
ADAM2 (uc003xnj.1) at chr8:39720412-39814936 - ADAM metallopeptidase domain 2 proprotein
ADAM28 (uc010lua.1) at chr8:24240012-24268671 - ADAM metallopeptidase domain 28 isoform 1
ADAM29 (uc010irr.1) at chr4:176133252-176135714 - ADAM metallopeptidase domain 29 preproprotein
ADAM20 (uc010ard.1) at chr14:70059047-70061377 - ADAM metallopeptidase domain 20 preproprotein
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 (uc003ujj.1) at chr7:87401638-87600049 - ADAM metallopeptidase domain 22 isoform 5
ADAM22 (uc003uji.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 (uc002vbq.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
YWHAB (uc002xmu.1) at chr20:42947758-42970575 - tyrosine 3-monooxygenase/tryptophan
YWHAB (uc002xmt.1) at chr20:42947758-42970575 - tyrosine 3-monooxygenase/tryptophan
YWHAE (uc002fsj.1) at chr17:1194593-1250267 - tyrosine 3/tryptophan 5 -monooxygenase
YWHAH (uc003alz.1) at chr22:30670479-30683590 - tyrosine 3-monooxygenase/tryptophan
YWHAQ (uc002qzx.1) at chr2:9641557-9688557 - tyrosine 3/tryptophan 5 -monooxygenase
YWHAQ (uc002qzw.1) at chr2:9641557-9688196 - tyrosine 3/tryptophan 5 -monooxygenase
YWHAZ (uc003yjx.1) at chr8:102000090-102034745 - tyrosine 3/tryptophan 5 -monooxygenase
YWHAZ (uc003yjw.1) at chr8:102000090-102033447 - tyrosine 3/tryptophan 5 -monooxygenase
YWHAZ (uc003yjv.1) at chr8:102000090-102032853 - tyrosine 3/tryptophan 5 -monooxygenase
YWHAZ (uc010mbr.1) at chr8:102000090-102034287 - tyrosine 3/tryptophan 5 -monooxygenase
ADAM21P (uc010arb.1) at chr14:69782223-69784271 - ADAM21-like protein.
RefSeq Genes
click
ADAM2 at chr8:39720412-39814936 - (NM_001464) ADAM metallopeptidase domain 2 proprotein
ADAM20 at chr14:70058831-70071485 - (NM_003814) ADAM metallopeptidase domain 20 preproprotein
ADAM21 at chr14:69993970-69996375 - (NM_003813) ADAM metallopeptidase domain 21 preproprotein
ADAM21P1 at chr14:69782223-69784271 - (NR_003951)
ADAM22 at chr7:87401638-87649364 - (NM_004194) ADAM metallopeptidase domain 22 isoform 4
ADAM22 at chr7:87401638-87664385 - (NM_021723) ADAM metallopeptidase domain 22 isoform 1
ADAM22 at chr7:87401638-87664385 - (NM_021722) ADAM metallopeptidase domain 22 isoform 2

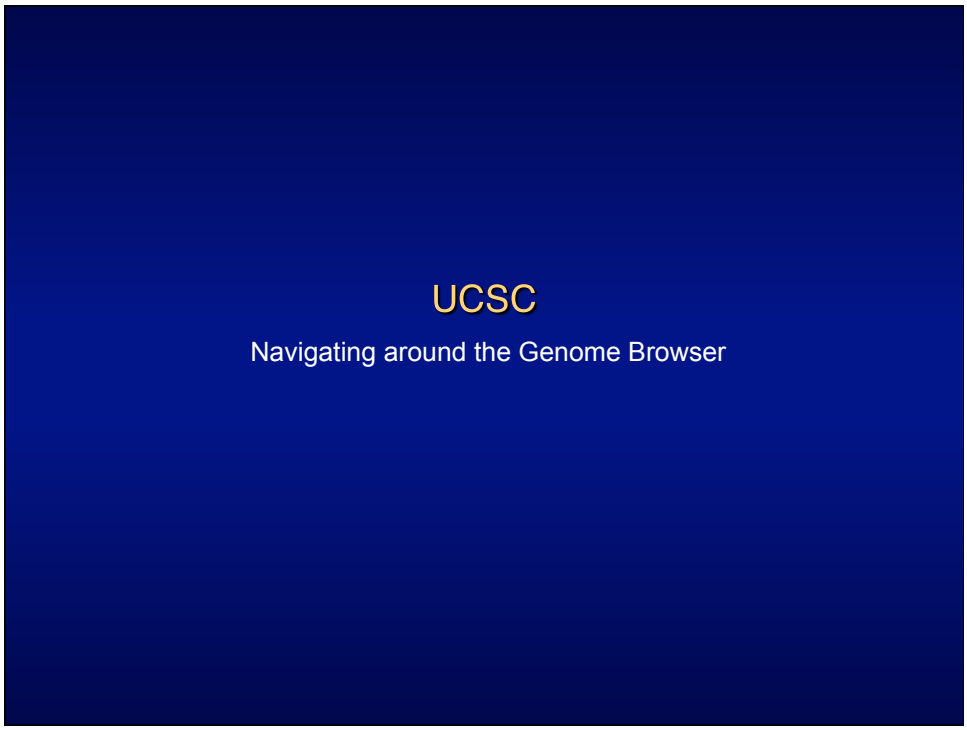

Human chr8:39,720,412-39,814,936 - UCSC Genome Browser v223
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help
UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)
intron
exon
Mapping and Sequencing Tracks
Base Position Chromosome Band STS Markers FISH Clones Recomb Rate Map Contigs

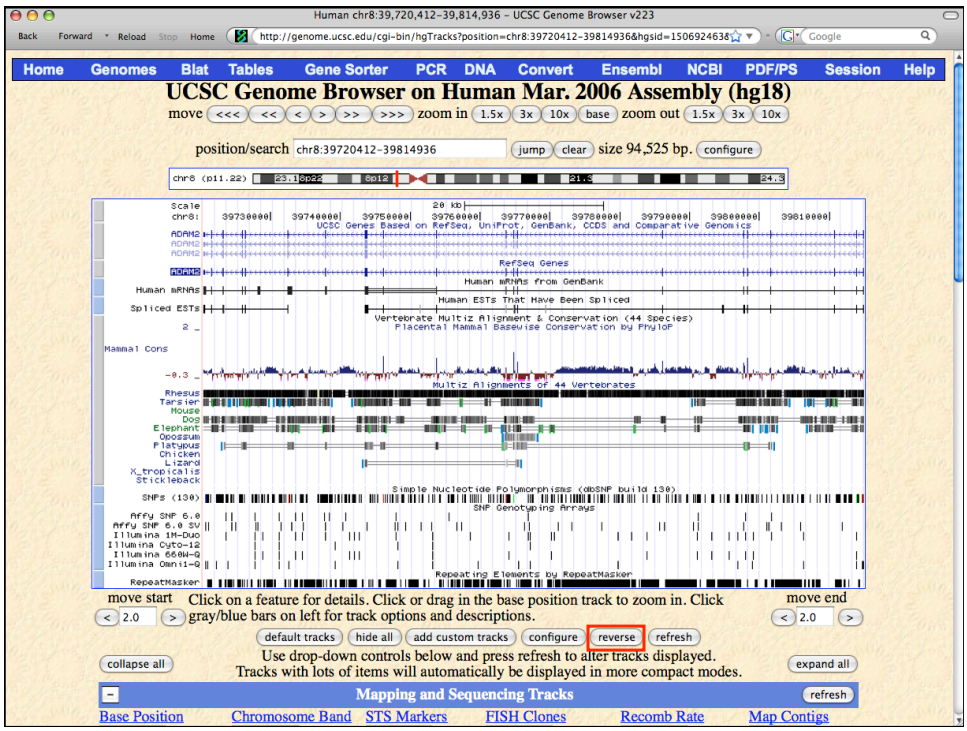

Human Gene ADAM2 (uc003xnj.1) Description and Page Index
UCSC Gene details
Human Gene ADAM2 (uc003xnj.1) Description and Page Index
Description: ADAM metallopeptidase domain 2 proprotein
RefSeq Summary (NM_001464): This gene encodes a member of the ADAM (a disintegrin and metalloprotease domain) family. 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. [provided by RefSeq].
Strand: - Genomic Size: 94525 Exon Count: 21 Coding Exon Count: 20
Page Index | Sequence and Links | UniProtKB Comments | CTD | Microarray | RNA Structure
Protein Structure | Other Species | GO Annotations | mRNA Descriptions | Other Names | Model Information
Methods
Sequence and Links to Tools and Databases
Genomic Sequence (chr8:39,720,412-39,814,936) | mRNA (may differ from genome) | Protein (735 aa)
Gene Sorter | Genome Browser | Protein FASTA | Proteome Browser | VisiGene | Table Schema
CGAP | Ensembl | Entrez Gene | ExonPrimer | GeneCards | GeneNetwork
Gepis Tissue | H-INV | HGNC | HPRD | Jackson Lab | OMIM
PubMed | Stanford SOURCE | Treefam | UniProtKB | User annotations
Comments and Description Text from UniProtKB
ID: ADAM2_HUMAN
DESCRIPTION: RecName: Full=Disintegrin and metalloproteinase domain-containing protein 2; Short=ADAM 2; AltName: Full=Fertilin subunit beta; AltName: Full=PH-30; Short=PH30; AltName: Full=PH30-beta; AltName: Full=Cancer/testis antigen 15; Short=CT15; Flags: Precursor;
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: Membrane; Single-pass 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.

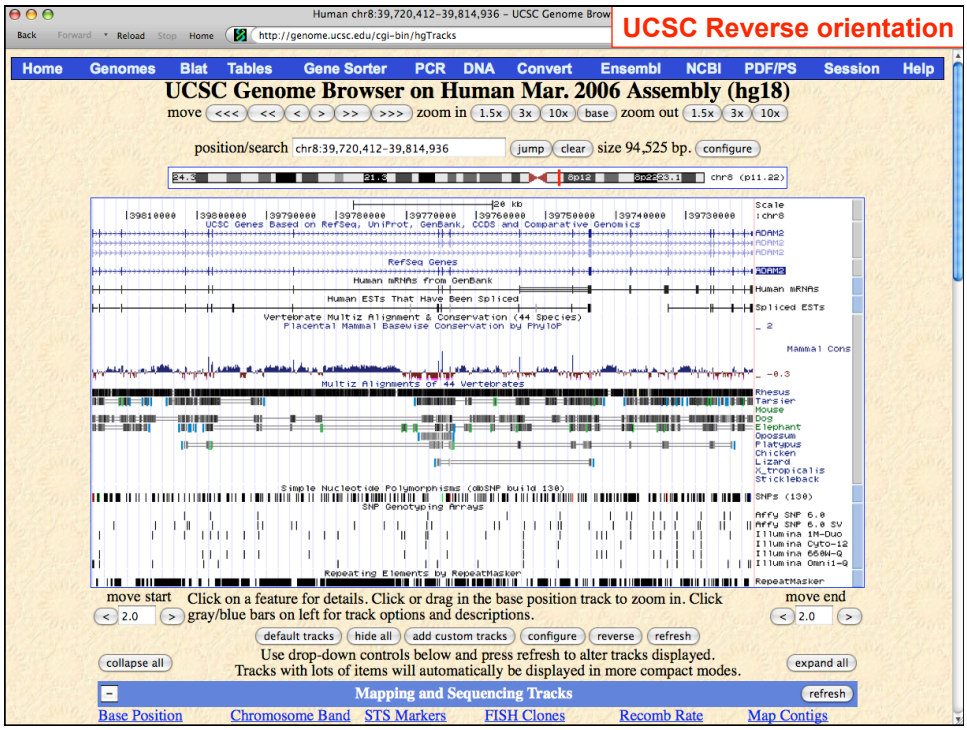

Human Gene ADAM2 (uc003xnj.1) Description and Page Index
Back Forward Reload Stop Home
http://genome.ucsc.edu/cgi-bin/hgGene?hgg_gene=uc003xnj.1&hgg_prot=Q99965&hgg_chro
UCSC Gene details
Affymetrix All Exon Microarrays
breast
cerebellum
heart
kidney
liver
muscle
pancreas
prostate
spleen
testes
thyroid
Ratios
mRNA Secondary Structure of 3' and 5' UTRs
Region Fold Energy Bases Energy/Base Display As
5' UTR -15.00 75 -0.200 Picture PostScript Text
3' UTR -72.02 359 -0.201 Picture PostScript Text
The RNAfold program from the Vienna RNA Package is used to perform the secondary structure predictions and folding calculations. The estimated folding energy is in kcal/mol. The more negative the energy, the more secondary structure the RNA is likely to have.
Protein Domain and Structure Information
InterPro Domains: Graphical view of domain structure
IPR006586 - ADAM_Cys-rich
IPR001762 - Blood-coag_inhib_Disintegrin
IPR018358 - Disintegrin_CS
IPR013032 - EGF-like_reg_CS
IPR013111 - EGF_extracell
IPR001590 - Peptidase_M12B
IPR002870 - Peptidase_M12B_N
Pfam Domains:
PF01421 - Reprolysin (M12B) family zinc metalloprotease
PF01562 - Reprolysin family propeptide
PF08516 - ADAM cysteine-rich
PF00200 - Disintegrin
PF07974 - EGF-like domain
SCOP Domains:

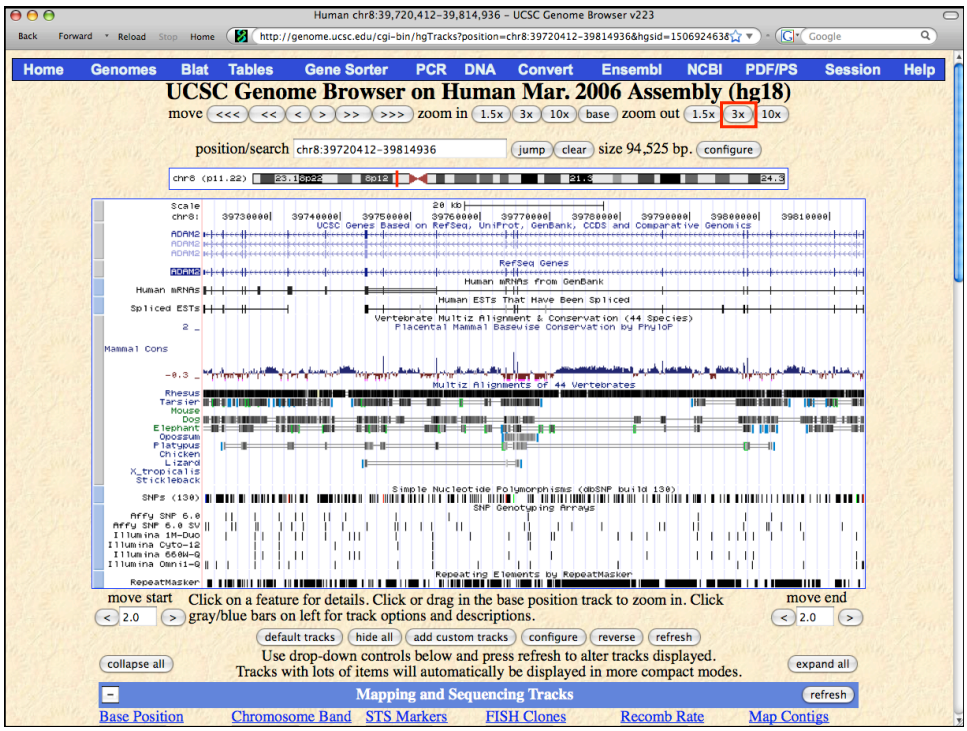

RefSeq Gene ADAM2
UCSC RefSeq Gene details
RefSeq: NM_001464.3 Status: Reviewed
Description: Homo sapiens ADAM metallopeptidase domain 2 (ADAM2), mRNA.
CCDS: CCDS34884.1
CDS: 3' complete
OMIM: 601533
Entrez Gene: 2515
PubMed on Gene: ADAM2
PubMed on Product: ADAM metallopeptidase domain 2 proprotein
GeneCards: ADAM2
AceView: ADAM2
Stanford SOURCE: NM_001464
CDS FASTA alignment from multiple alignment: 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. [provided by RefSeq].
mRNA/Genomic Alignments
SIZE IDENTITY CHROMOSOME STRAND START END QUERY START END TOTAL
2642 100.0% 8 - 39720412 39814936 NM_001464 1 2642 2657
Links to sequence:
Predicted Protein
mRNA Sequence
Genomic Sequence
click
different from the genomic sequence.
from assembly

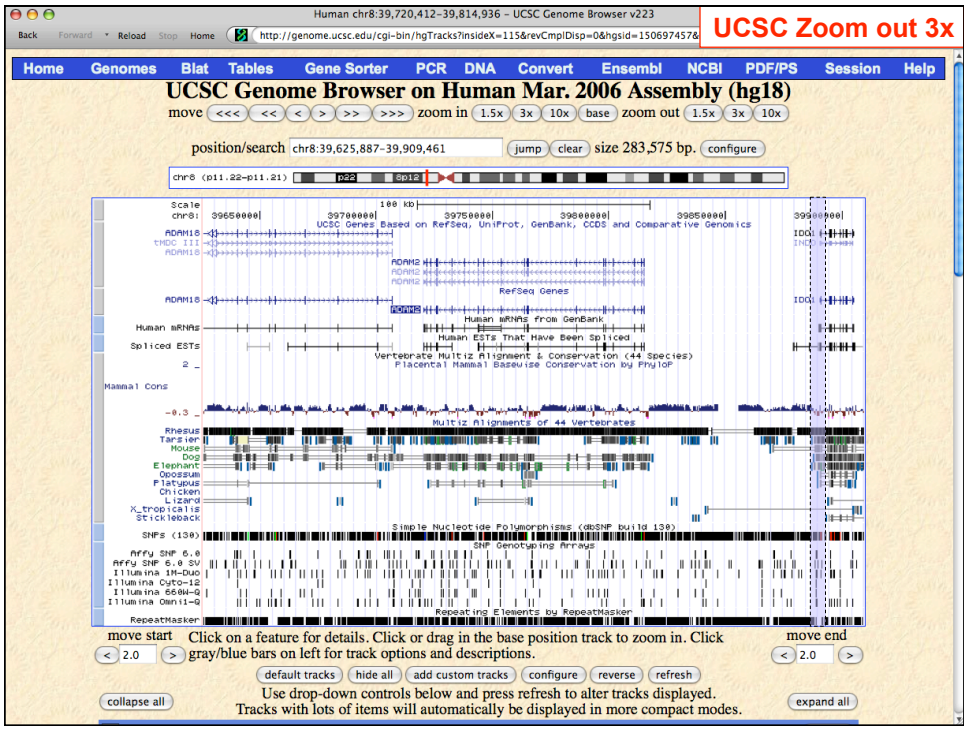

Genomic Sequence Near Gene
Back Forward Reload Stop Home
http://genome.ucsc.edu/cgi-bin/hgc?hgsid=150692463&g=htcGeneInGe
UCSC RefSeq Gene details
Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help
Genomic Sequence Near Gene
Get Genomic Sequence Near Gene
Note: if you would prefer 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
Split UTR and CDS parts of an exon into separate FA
Note: if a feature is close to the beginning or end of a chro
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
ggaagtatctaccaacacatacccctgtgatccgacaacttcactctagaa
atatacacagtagaaatccttacttattacaccaaaaggcatgtagaaga
atgtttatagctaaatattatttttaatagctggaacaataaacaaacaa
aatattcattaacagtaaaatggaaacacaagtggtgttatatttaatga
atttgtaatatacaccaatgaggataaaacagaactattgctttgtatga
accttacaatcatcctattaaaagaacccagacatgaaagagtagatggt
gattgcttctacttgcgaaaagttcaaaaacagacaaaacgaatctttgg
ttgtttagaagtcatggttgtgaggttggaatcctgggatttgggtggtt
ctttttcatttcttcatcgtggtactagttacgctgtttttttttccac
ttgaatattaattgacctgtgaacttatgatttattatatactttttttc
gtttttttgtttcgttctttttttttttttttttttttttgacggaat
ttcgctctctcacccaggctggagtgcagtggtaaggtctttctgctcac
tgcaaccctctgcctcctaggttcaagcgattctcctgcctcagcttcccg
agtagctgggatttcaggcacccgccaccatgcctggctaattttttttt
gtatttttagtacagacggggtttcaccatgttgcggaaactggtctcga
actcctgatacgtgtttatatatattcaattgaaattttacttaagaagt
gtttataaatcttcctgttcctcagctgttggaagtgattttgttgttgc
tgttgccttaattaggcatcacgtccagtggagtgcttgtcctgcaagag
acagggctcaggagctcgcacgttccacagcgcaccacacccaacctcag
cccacctgggctctcccagccgcctacctcttccaggctgcgtggccggg

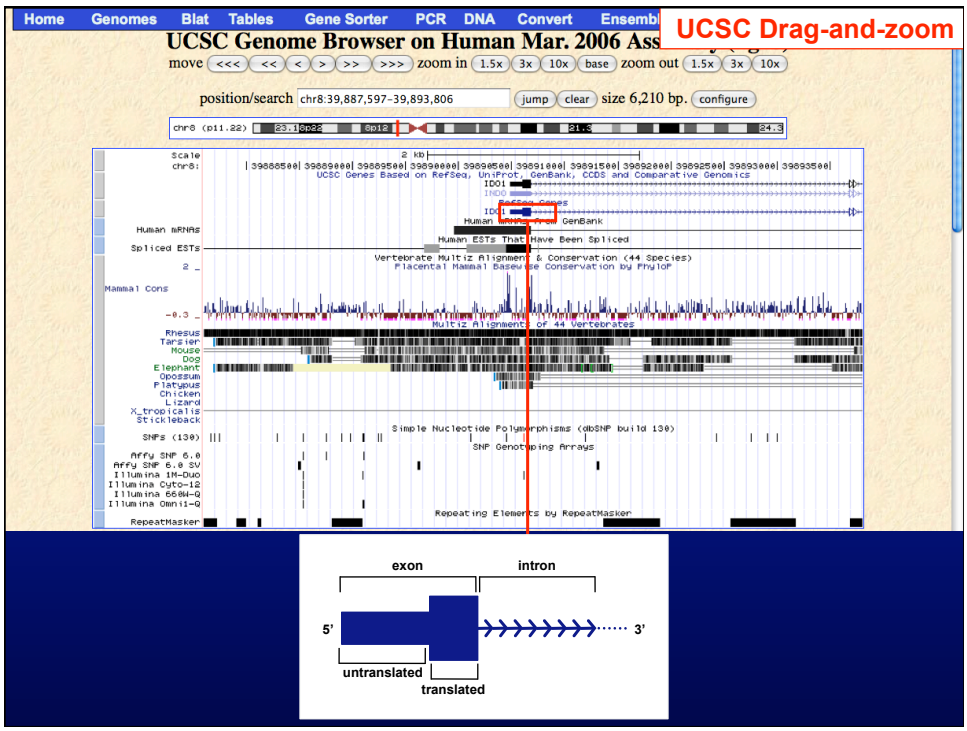

UCSC
Navigating around the Genome Browser


Human chr8:39,720,412-39,814,936 - UCSC Genome Browser v223
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help
UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)
move zoom in zoom out
position/search chr8:39720412-39814936 jump clear size 94,525 bp. configure
default tracks hide all add custom tracks configure reverse refresh
collapse all expand all
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 refresh
Base Position Chromosome Band STS Markers FISH Clones Recomb Rate Map Contigs


UCSC Reverse orientation
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help
UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)
move zoom in zoom out
position/search chr8:39,720,412-39,814,936 jump clear size 94,525 bp. configure
default tracks hide all add custom tracks configure reverse refresh
collapse all expand all
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 refresh
Base Position Chromosome Band STS Markers FISH Clones Recomb Rate Map Contigs

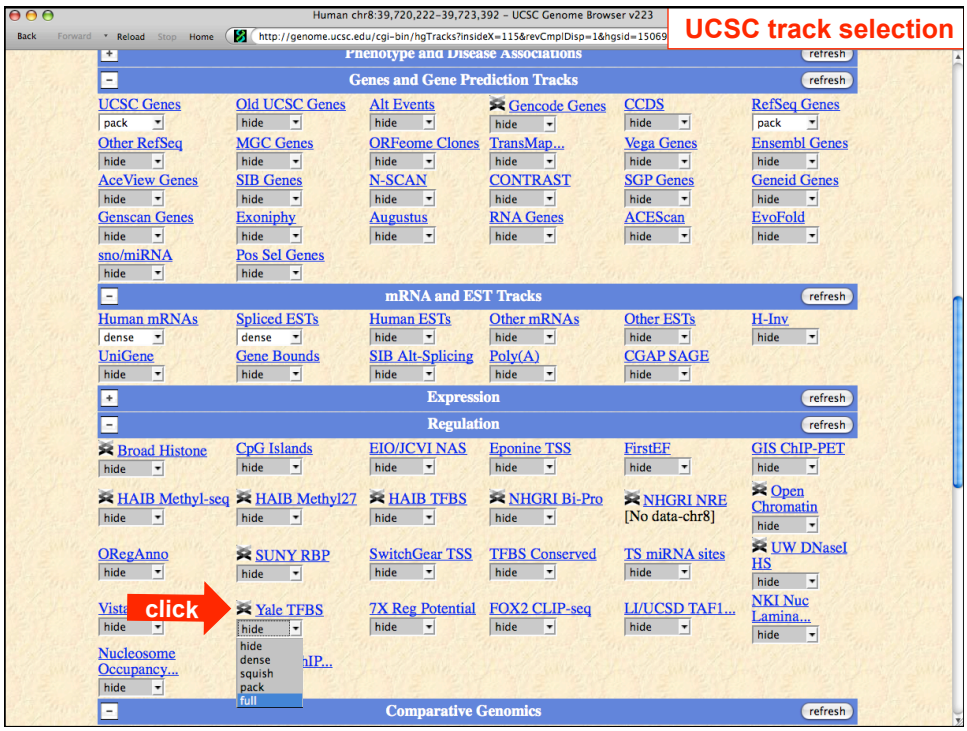

Human chr8:39,720,412-39,814,936 - UCSC Genome Browser v223
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help
UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chr8:39720412-39814936 jump clear size 94,525 bp. configure
Mapping and Sequencing Tracks
Base Position Chromosome Band STS Markers FISH Clones Recomb Rate Map Contigs

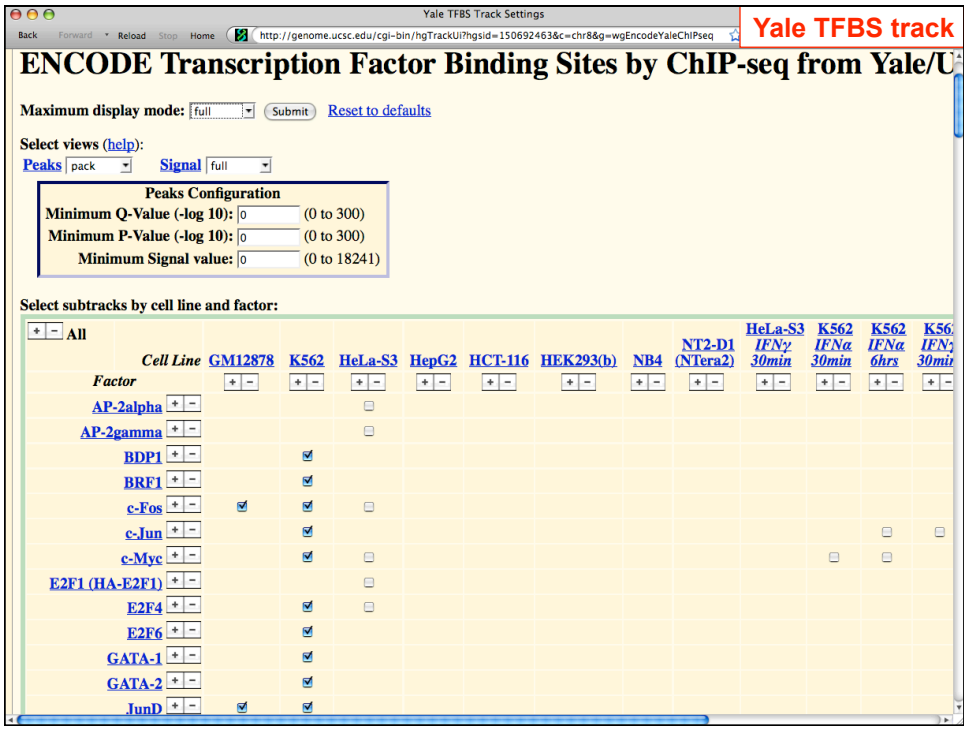

UCSC Zoom out 3x
Human chr8:39,720,412-39,814,936 - UCSC Genome Browser v223
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help
UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chr8:39,625,887-39,909,461 jump clear size 283,575 bp. configure

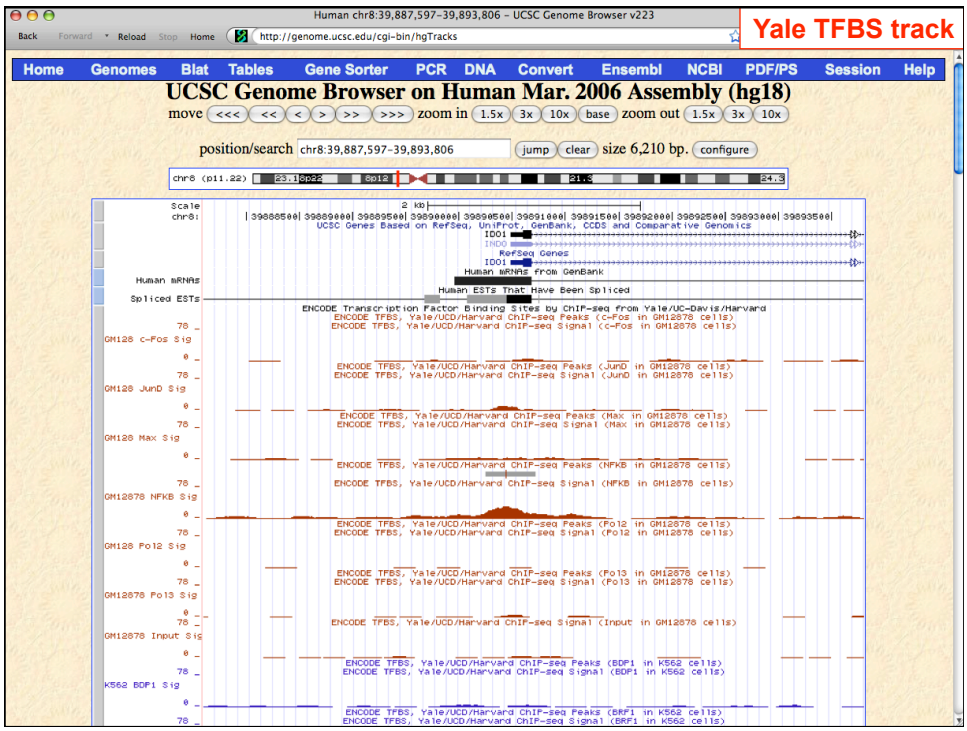


UCSC

Add a track to the Genome Browser


UCSC track selection
Human chr8:39,720,222-39,723,392 - UCSC Genome Browser v223
Genes and Gene Prediction Tracks
mRNA and EST Tracks
Expression
Regulation
Comparative Genomics
click
Yale TFBS
hide
dense
squish
pack
full

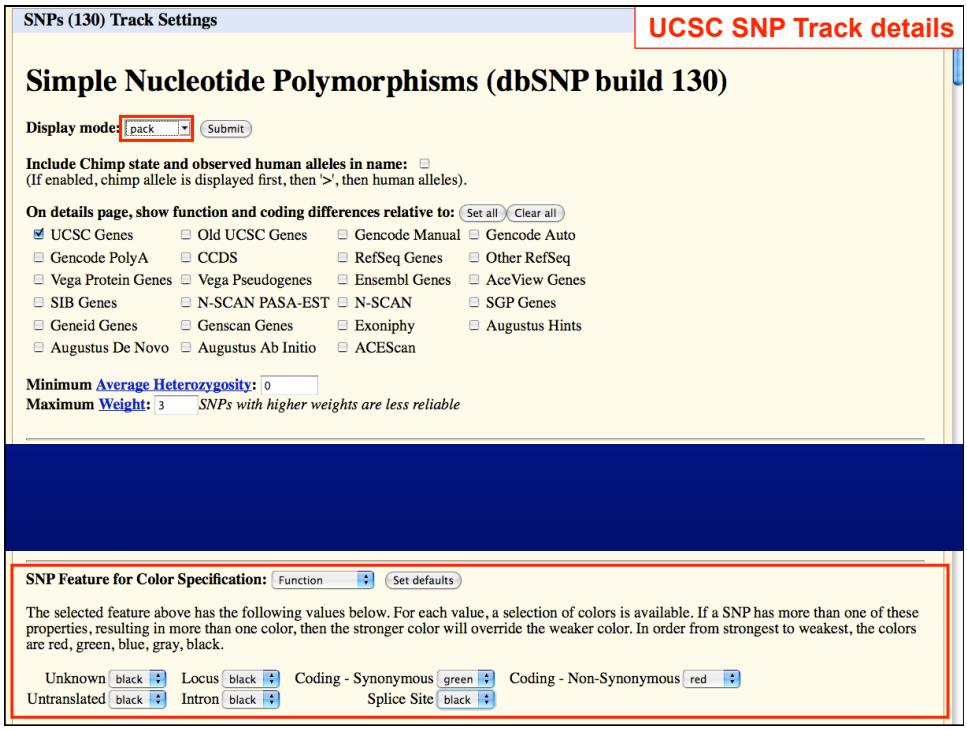

Yale TFBS track
Yale TFBS Track Settings
ENCODE Transcription Factor Binding Sites by ChIP-seq from Yale/U
Maximum display mode: full
Submit
Reset to defaults
Select views (help):
Peaks pack
Signal full
Peaks Configuration
Minimum Q-Value (-log 10): 0 (0 to 300)
Minimum P-Value (-log 10): 0 (0 to 300)
Minimum Signal value: 0 (0 to 18241)
Select subtracks by cell line and factor:
All
Cell Line
Factor
GM12878
K562
HeLa-S3
HepG2
HCT-116
HEK293(b)
NB4
NT2-D1 (NTera2)
HeLa-S3 IFNγ 30min
K562 IFNα 30min
K562 IFNα 6hrs
AP-2alpha
AP-2gamma
BDP1
BRF1
c-Fos
c-Jun
c-Myc
E2F1 (HA-E2F1)
E2F4
E2F6
GATA-1
GATA-2
JunD

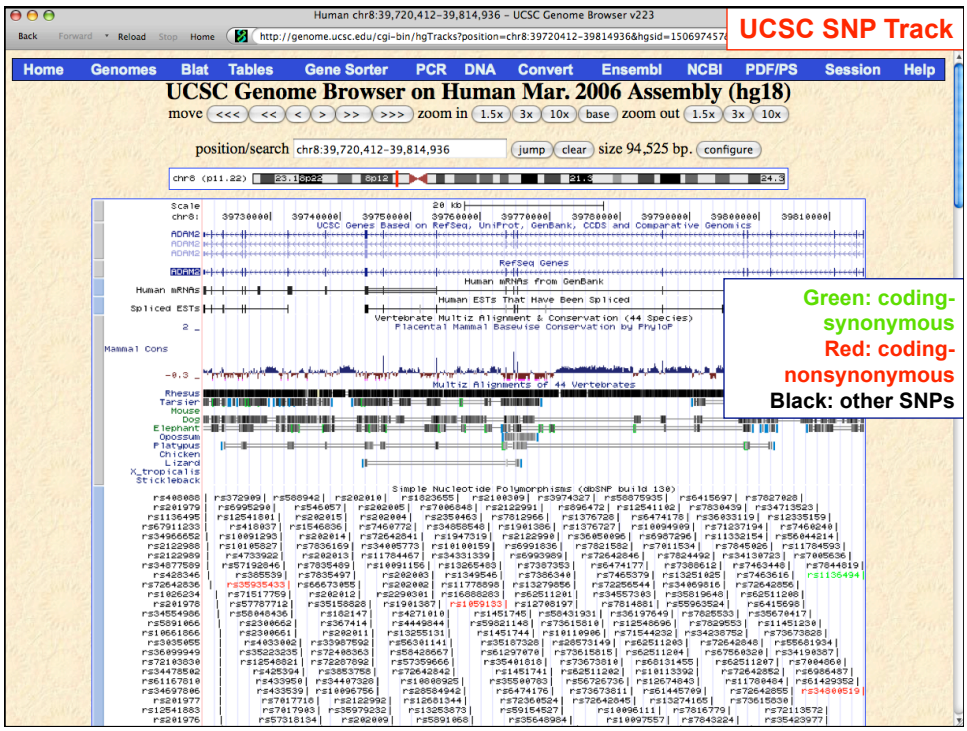


UCSC

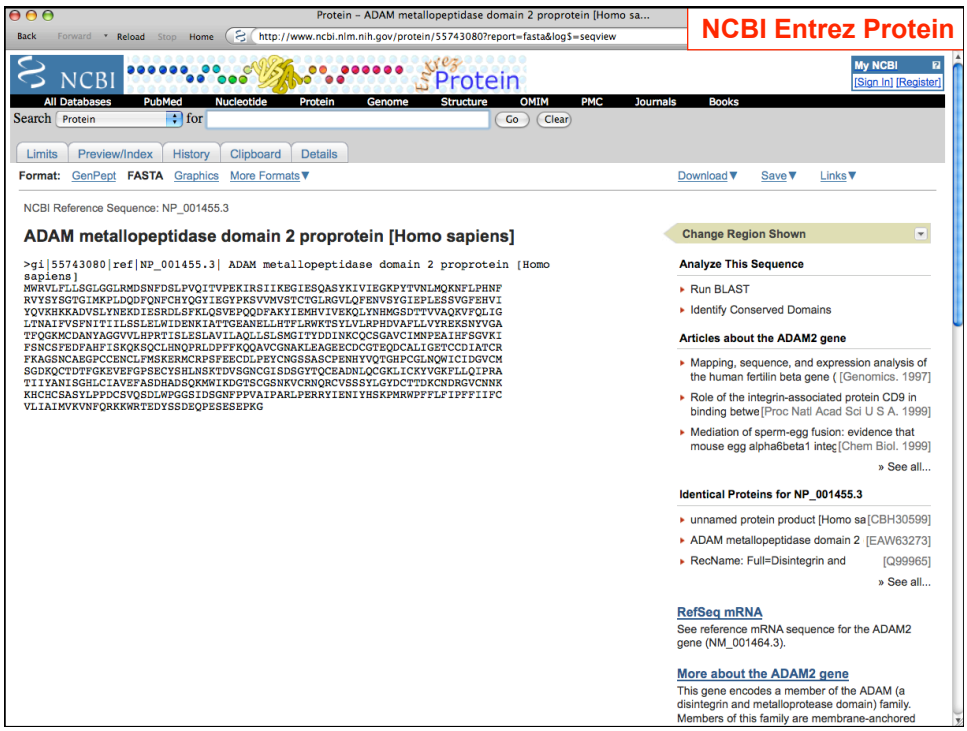
Change the color of items in a track

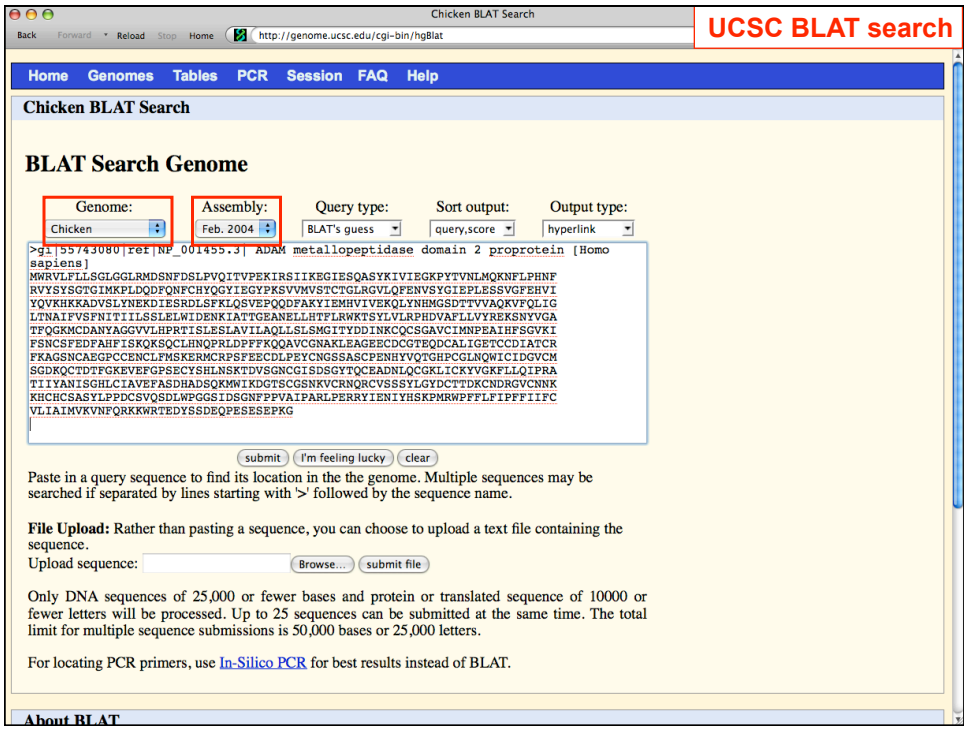

Human chr8:39,887,597-39,893,806 - UCSC Genome Browser v223
http://genome.ucsc.edu/cgi-bin/hgTracks
Lamprey Net
Medaka Chain
Medaka Net
Stickleback Chain
Stickleback Net
Fugu Chain
Fugu Net
Tetraodon Chain
Tetraodon Net
Tetraodon Ecores
Zebrafish Chain
Zebrafish Net
X. tropicalis Chain
X. tropicalis Net
Zebra finch Chain
Zebra finch Net
Lizard Chain
Lizard Net
Platypus Chain
Platypus Net
Chimp Chain/Net
Orangutan Chain/Net
Rhesus Chain/Net
Guinea Pig Chain
Guinea Pig Net
Rat Chain
Rat Net
Marmoset Chain/Net
Mouse Chain
Mouse Net
Dog Chain/Net
Cat Chain/Net
Horse Chain/Net
Cow Chain/Net
Opossum Chain/Net
Chicken Chain/Net
Variation and Repeats
refresh
click
SNPs (130)
SNPs (129)
SNPs (128)
SNPs (126)
SNP Arrays
HGDP Allele Freq
HGDP Smoothd FST
HGDP Hetrzygsty
HGDP iHS
HGDP XP-EHH
HapMap SNPs
HapMap LD Phased
Tajima's D SNPs
Tajima's D
DGV Struct Var
HGSV Discordant
Segmental Dups
Structural Var
Exapted Repeats
RepeatMasker
RepMask 3.2.7
Interrupted Rpts
Intr Rpts 3.2.7
Simple Repeats
Microsatellite
Self Chain
Genome Variants
Pilot ENCODE Regions and Genes
Pilot ENCODE Transcription
Pilot ENCODE Chromatin Immunoprecipitation
Pilot ENCODE Chromatin Structure
Pilot ENCODE Comparative Genomics and Variation

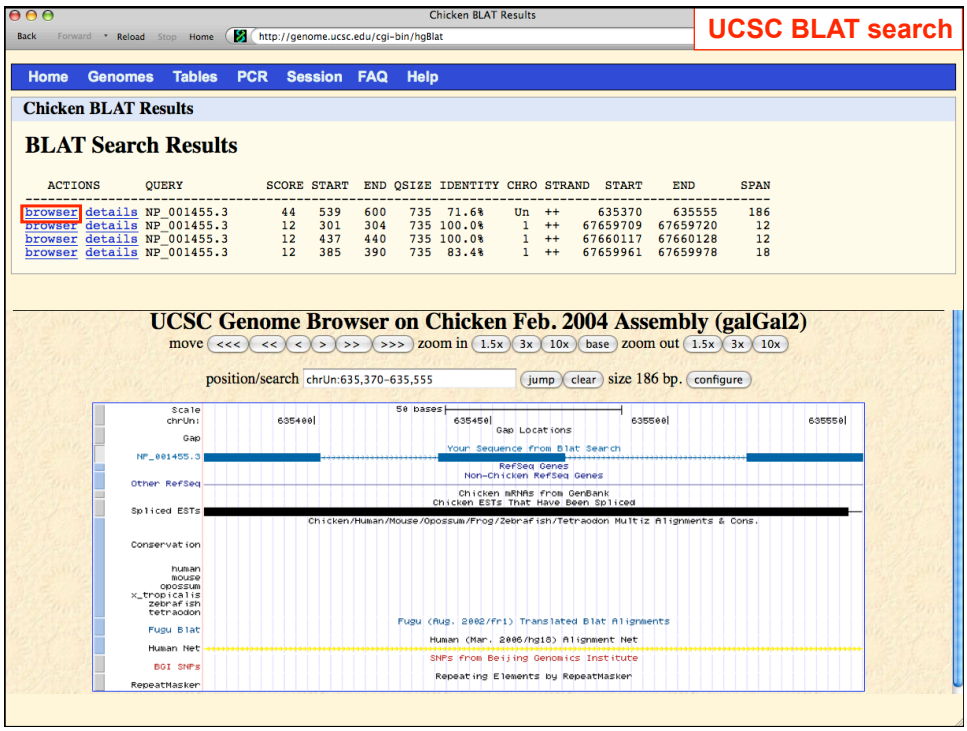

SNPs (130) Track Settings
UCSC SNP Track details
Simple Nucleotide Polymorphisms (dbSNP build 130)
Display mode: pack
Submit
Include Chimp state and observed human alleles in name:
(If enabled, chimp allele is displayed first, then '>', then human alleles).
On details page, show function and coding differences relative to:
Set all
Clear all
UCSC Genes
Old UCSC Genes
Gencode Manual
Gencode Auto
Gencode PolyA
CCDS
RefSeq Genes
Other RefSeq
Vega Protein Genes
Vega Pseudogenes
Ensembl Genes
AceView Genes
SIB Genes
N-SCAN PASA-EST
N-SCAN
SGP Genes
Geneid Genes
Genscan Genes
Exoniphy
Augustus Hints
Augustus De Novo
Augustus Ab Initio
ACEScan
Minimum Average Heterozygosity: 0
Maximum Weight: 3
SNPs with higher weights are less reliable
SNP Feature for Color Specification: Function
Set defaults
The selected feature above has the following values below. For each value, a selection of colors is available. If a SNP has more than one of these properties, resulting in more than one color, then the stronger color will override the weaker color. In order from strongest to weakest, the colors are red, green, blue, gray, black.
Unknown black
Locus black
Coding - Synonymous green
Coding - Non-Synonymous red
Untranslated black
Intron black
Splice Site black

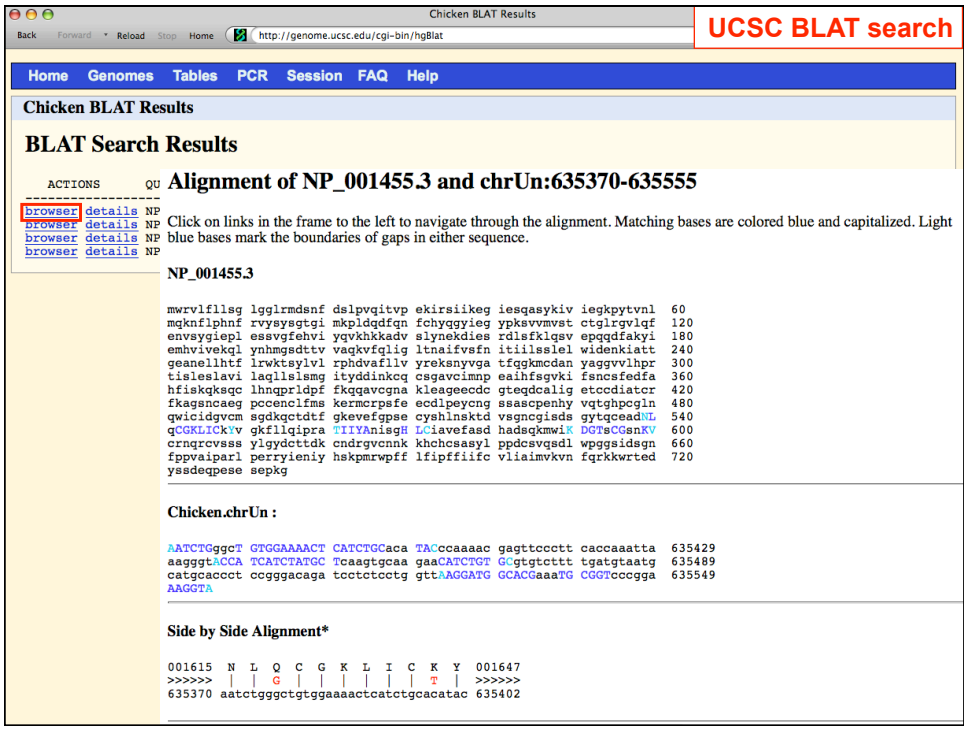

UCSC SNP Track
Human chr8:39,720,412-39,814,936 - UCSC Genome Browser v223
Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help
UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)
position/search chr8:39,720,412-39,814,936
Green: coding-synonymous
Red: coding-nonsynonymous
Black: other SNPs


UCSC
Find a chicken homolog of a human protein


NCBI Entrez Protein
Protein – ADAM metallopeptidase domain 2 proprotein [Homo sa...
http://www.ncbi.nlm.nih.gov/protein/55743080?report=fasta&log$=seqview
NCBI
Protein
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Protein for Go Clear
Limits Preview/Index History Clipboard Details
Format: GenPept FASTA Graphics More Formats▼
Download▼ Save▼ Links▼
NCBI Reference Sequence: NP_001455.3
ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
>gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domain 2 proprotein [Homo sapiens]
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKIVIEGKPYTVNLMQKNFLPHNF
RVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPLESSVGFEHVI
YQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDFAKYIEMHVIVEKQLYNHMGSDTTVVAQKVFQLIG
LTNAIFVSFNITIILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLRPHDVAFLLVYREKSNYVGA
TFQGKMCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAIHFSGVKI
FSNCSFEDFAHFISKQKSQCLHNQPRLDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCM
SGDKQCTDTFGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYTQCEADNLQCGKLICKYVGKFLLQIPRA
TIIYANISGHLCIAVEFASDHADSQKMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTDKCNDRGVCNNK
KHCHCSASYLPPDCSVQSDLWPGGSIDSGNFPPVAIPARLPERRYIENIYHSKPMRWPFFLFIPFFIIFC
VLIAIMVKVNFQRKKWRTEDYSSDEQPESESEPKG
Change Region Shown
Analyze This Sequence
Run BLAST
Identify Conserved Domains
Articles about the ADAM2 gene
Mapping, sequence, and expression analysis of the human fertilin beta gene ([Genomics. 1997]
Role of the integrin-associated protein CD9 in binding betwe [Proc Natl Acad Sci U S A. 1999]
Mediation of sperm-egg fusion: evidence that mouse egg alpha6beta1 intec [Chem Biol. 1999]
» See all...
Identical Proteins for NP_001455.3
unnamed protein product [Homo sa [CBH30599]
ADAM metallopeptidase domain 2 [EAW63273]
RecName: Full=Disintegrin and [Q99965]
» See all...
RefSeq mRNA
See reference mRNA sequence for the ADAM2 gene (NM_001464.3).
More about the ADAM2 gene
This gene encodes a member of the ADAM (a disintegrin and metalloprotease domain) family. Members of this family are membrane-anchored


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
YQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDFAKYIEMHVIVEKQLYNHMGSDTTVVAQKVFQLIG
LTNAIFVSFNITIILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLRPHDVAFLLVYREKSNYVGA
TFQGKMCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAIHFSGVKI
FSNCSFEDFAHFISKQKSQCLHNQPRLDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCM
SGDKQCTDTFGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYTQCEADNLQCGKLICKYVGKFLLQIPRA
TIIYANISGHLCIAVEFASDHADSQKMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTDKCNDRGVCNNK
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


Chicken BLAT Results
http://genome.ucsc.edu/cgi-bin/hgBlat
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 (galGal2)
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
Gap Locations
Your Sequence from Blat Search
RefSeq Genes
Non-Chicken RefSeq Genes
Chicken mRNAs from GenBank
Chicken ESTs That Have Been Spliced
Chicken/Human/Mouse/Opossum/Frog/Zebrafish/Tetraodon Multiz Alignments & Cons.
Fugu (Aug. 2002/fr1) Translated Blat Alignments
Human (Mar. 2006/hg18) Alignment Net
SNPs from Beijing Genomics Institute
Repeating Elements by RepeatMasker


Chicken BLAT Results
http://genome.ucsc.edu/cgi-bin/hgBlat
UCSC BLAT search
Home Genomes Tables PCR Session FAQ Help
Chicken BLAT Results
BLAT Search Results
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 fchyggyieg ypksvvmvst ctglrgvlqf 120
envsygiepl essvgfehvi yqvkhkkadv slynekdies rdlsfklqsv epqqdfakyi 180
emhvivekql ynhmgsdttv vaqkvfqlig ltnaifvsfn itiilsslel widenkiatt 240
geanellhtf lrwktsylvl rphdvafllv yreksnyvga tfqgkmcdan yagqvvlhpr 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
635370 aatctgggctgtggaaaactcatctgcacatac 635402

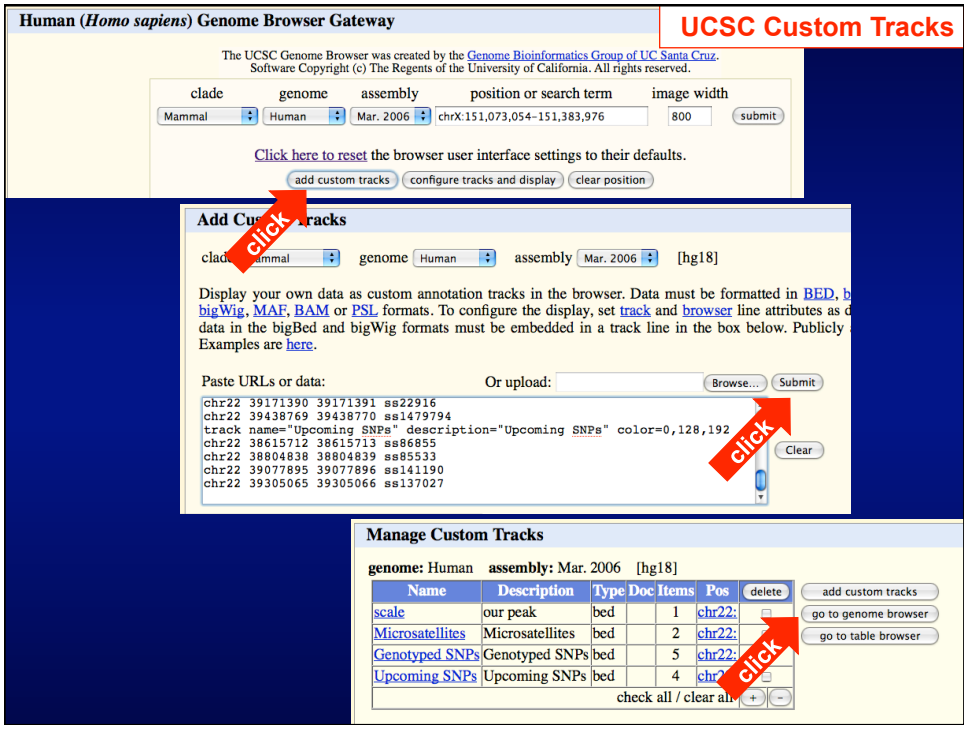
UCSC

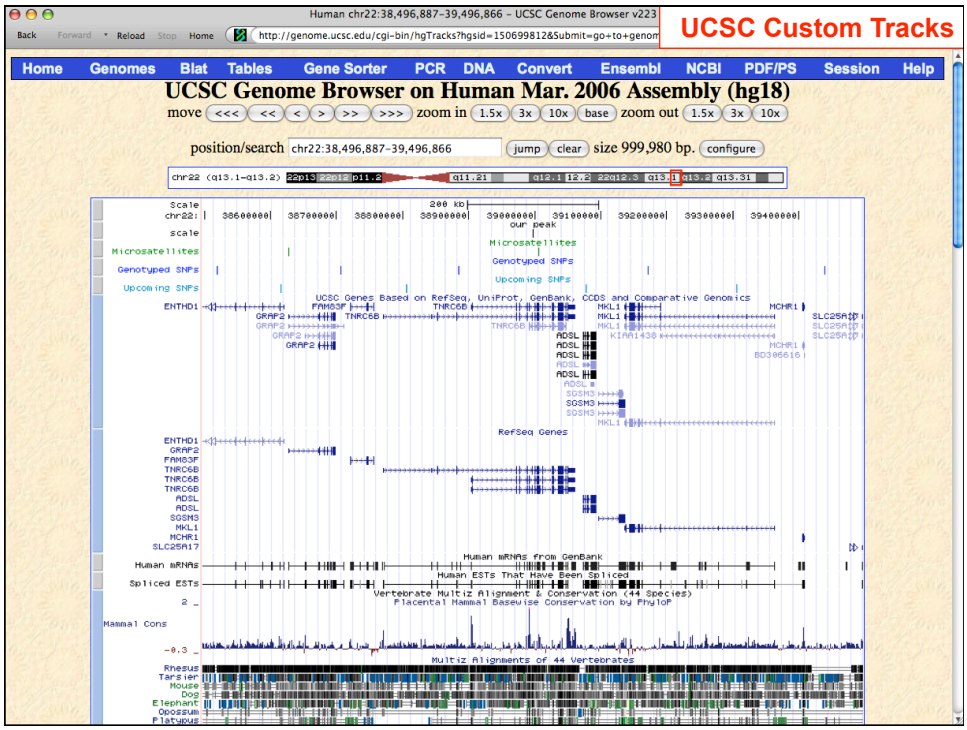
Add your own custom tracks

UCSC custom track format

```
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
chr22 38705963 38705964 ss2941443
chr22 38884157 38884158 ss141110
chr22 39171390 39171391 ss22916
chr22 39438769 39438770 ss1479794
track name="Upcoming SNPs" description="Upcoming SNPs" color=0,128,192
chr22 38615712 38615713 ss86855
chr22 38804838 38804839 ss85533
chr22 39077895 39077896 ss141190
chr22 39305065 39305066 ss137027
```

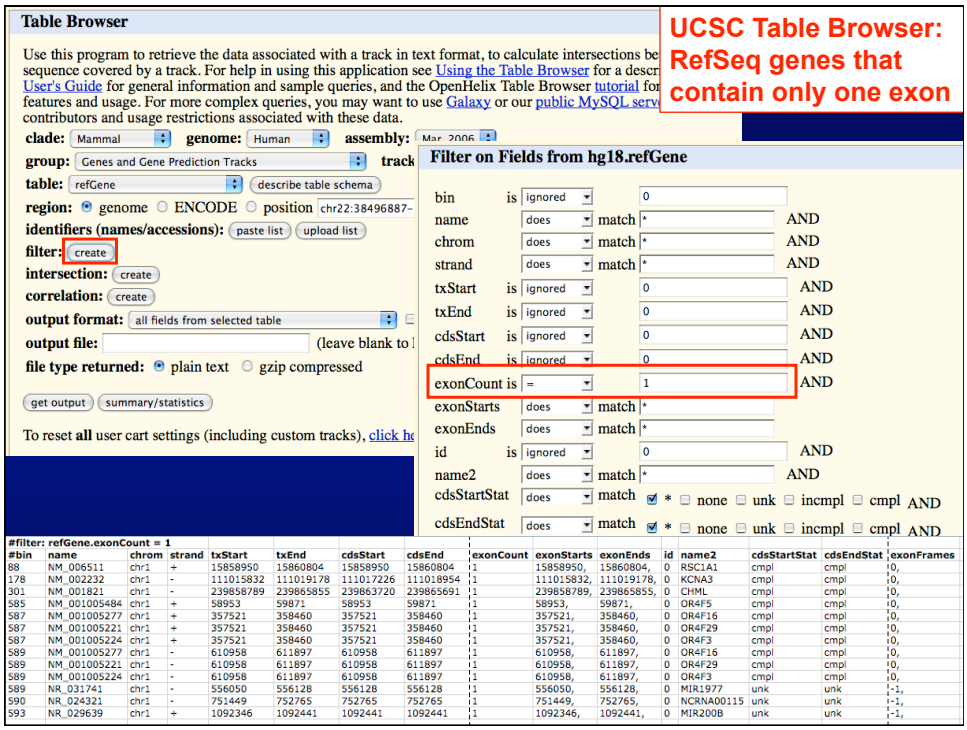
http://research.nhgri.nih.gov/teaching/custom_tracks.shtml


Human (*Homo sapiens*) Genome Browser Gateway
UCSC Custom Tracks
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
Mammal
Human
Mar. 2006
chrX:151,073,054-151,383,976
800
submit
Click here to reset the browser user interface settings to their defaults.
add custom tracks
configure tracks and display
clear position
click
Add Custom Tracks
genome Human
assembly Mar. 2006 [hg18]
Display your own data as custom annotation tracks in the browser. Data must be formatted in BED, bigWig, MAF, BAM or PSL formats. To configure the display, set track and browser line attributes as data in the bigBed and bigWig formats must be embedded in a track line in the box below. Publicly Examples are here.
Paste URLs or data:
Or upload:
Browse...
Submit
chr22 39171390 39171391 ss22916
chr22 39438769 39438770 ss1479794
track name="Upcoming SNPs" description="Upcoming SNPs" color=0,128,192
chr22 38615712 38615713 ss86855
chr22 38804838 38804839 ss85533
chr22 39077895 39077896 ss141190
chr22 39305065 39305066 ss137027
click
Clear
Manage Custom Tracks
genome: Human assembly: Mar. 2006 [hg18]
Name
Description
Type
Doc
Items
Pos
delete
scale
our peak
bed
1
chr22:
Microsatellites
Microsatellites
bed
2
chr22:
Genotyped SNPs
Genotyped SNPs
bed
5
chr22:
Upcoming SNPs
Upcoming SNPs
bed
4
chr2
check all / clear all
add custom tracks
go to genome browser
go to table browser
click


Human chr22:38,496,887-39,496,866 - UCSC Genome Browser v223
UCSC Custom Tracks
Home
Genomes
Blat
Tables
Gene Sorter
PCR
DNA
Convert
Ensembl
NCBI
PDF/PS
Session
Help
UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)
move
zoom in
zoom out
position/search
chr22:38,496,887-39,496,866
jump
clear
size 999,980 bp.
configure
Scale
scale
Microsatellites
Genotyped SNPs
Upcoming SNPs
our peak
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
RefSeq Genes
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
Vertebrate Multiz Alignment & Conservation (44 Species)
Placental Mammal Basewise Conservation by PhyloP
Mammal Cons
Multiz Alignments of 44 Vertebrates
Rhesus
Tarsier
Mouse
Dog
Elephant
Opossum
Platypus

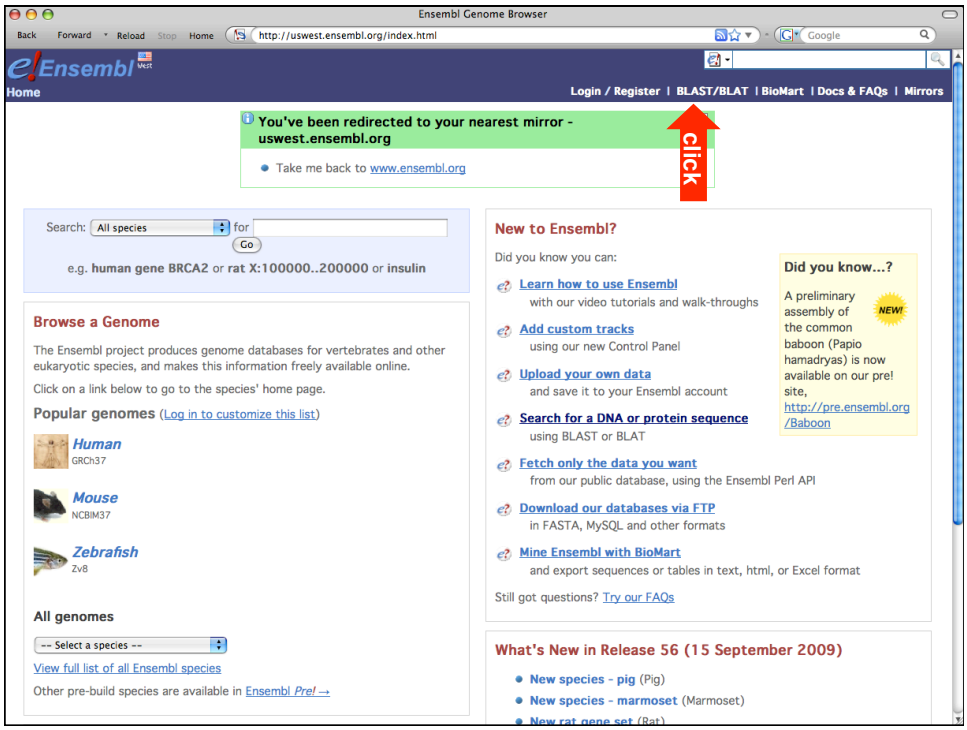
UCSC Table Browser

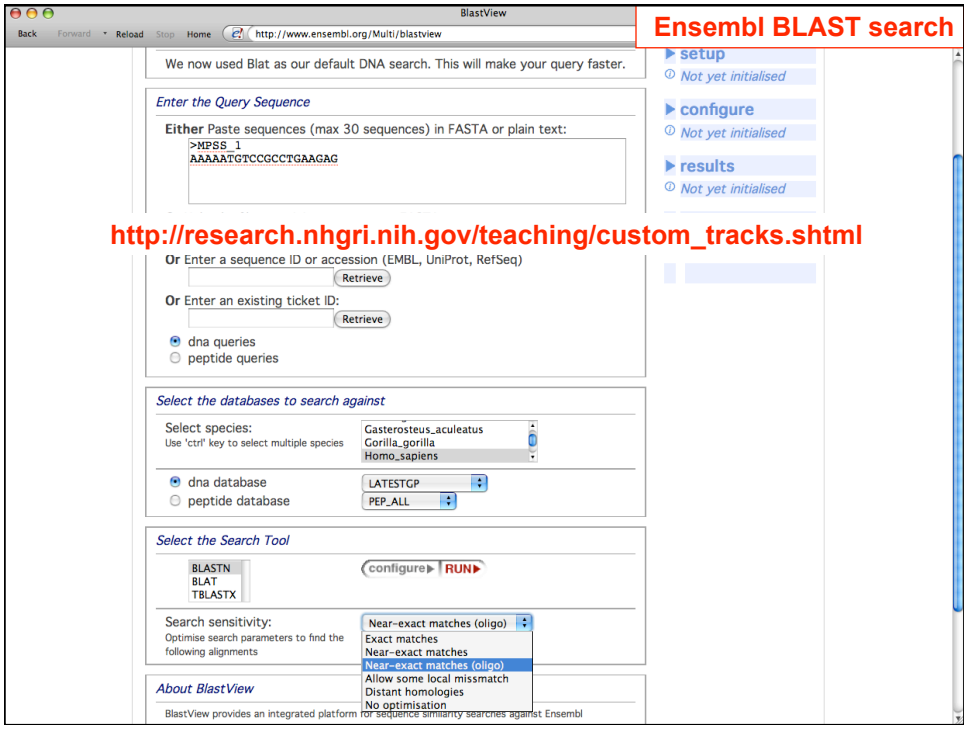
- Download track in text format
- Retrieve DNA sequence covered by a track
- Calculate intersections between tracks and view in the Genome Browser.
  - List all SNPs in a gene
- Filter track data based on certain criteria
  - Show all RefSeq genes that contain only one exon

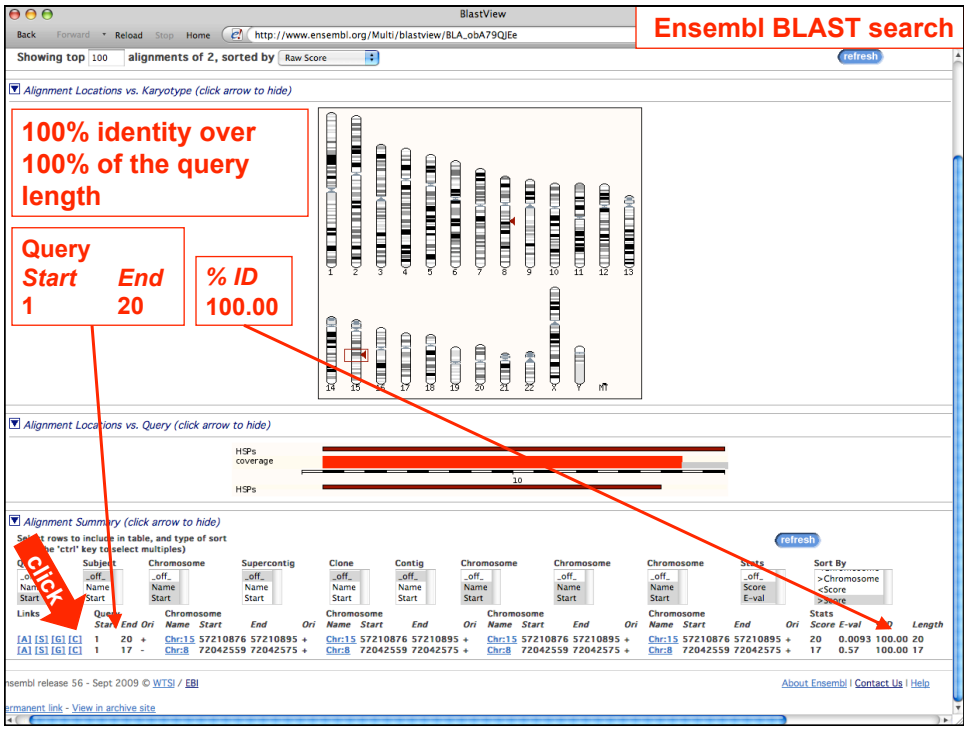


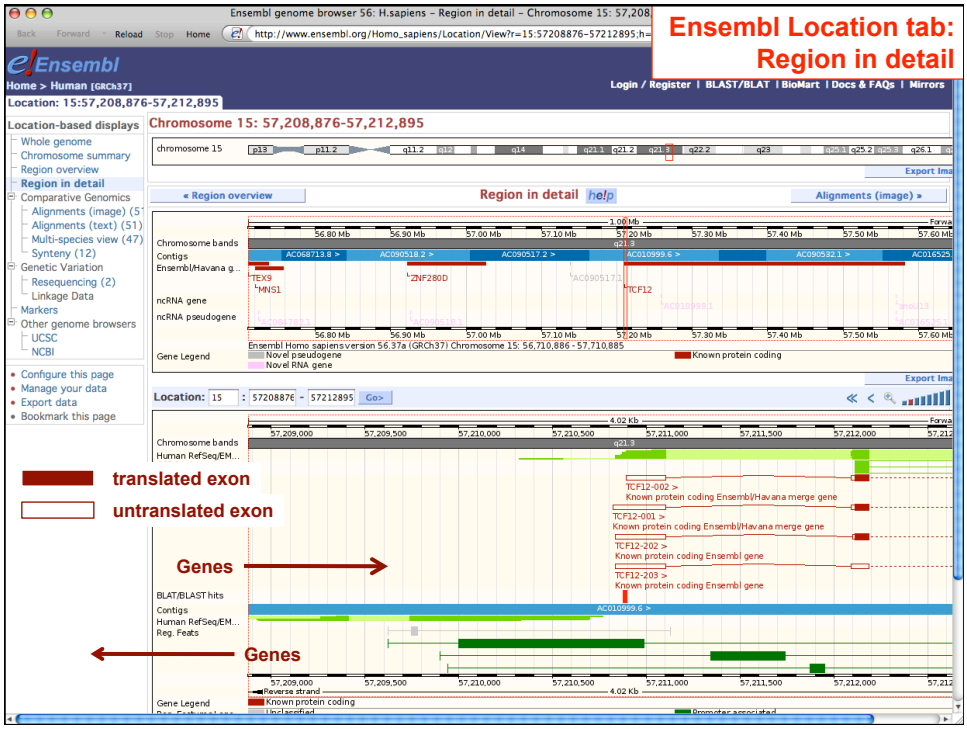
Ensembl

Identify genes that overlap with an oligo tag



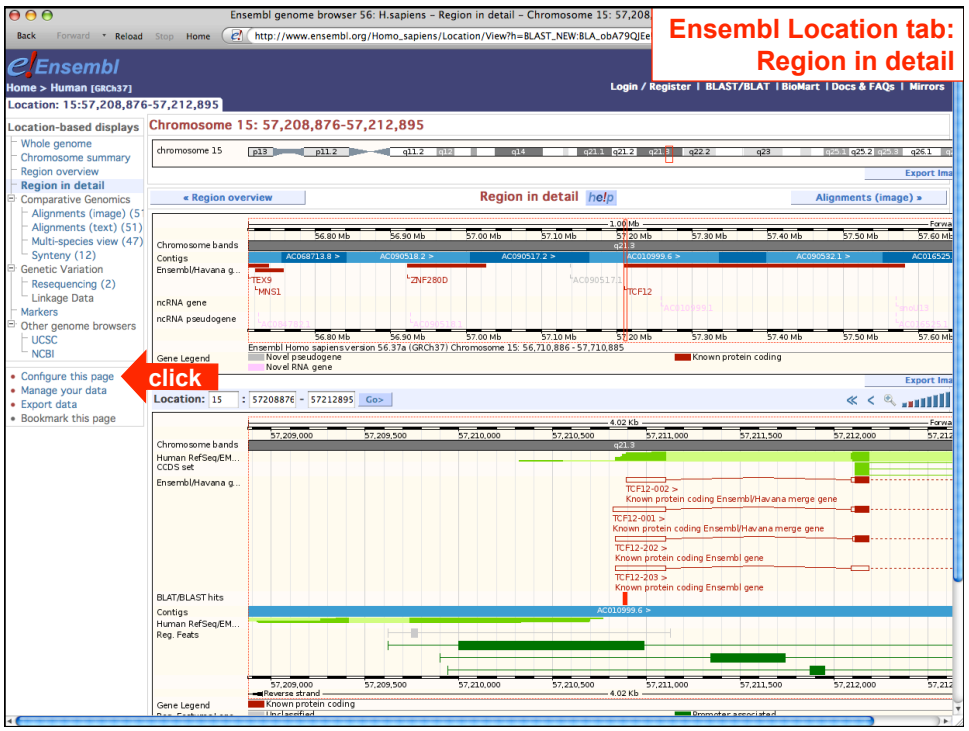

Ensembl BLAST search
BlastView
http://www.ensembl.org/Multi/blastview
We now used Blat as our default DNA search. This will make your query faster.
Enter the Query Sequence
Either Paste sequences (max 30 sequences) in FASTA or plain text:
>MPSS_1
AAAAATGTCCGCCTGAAGAG
setup
Not yet initialised
configure
Not yet initialised
results
Not yet initialised
http://research.nhgri.nih.gov/teaching/custom_tracks.shtml
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
Gasterosteus_aculeatus
Gorilla_gorilla
Homo_sapiens
dna database
peptide database
LATESTGP
PEP_ALL
Select the Search Tool
BLASTN
BLAT
TBLASTX
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
BlastView provides an integrated platform for sequence similarity searches against Ensembl

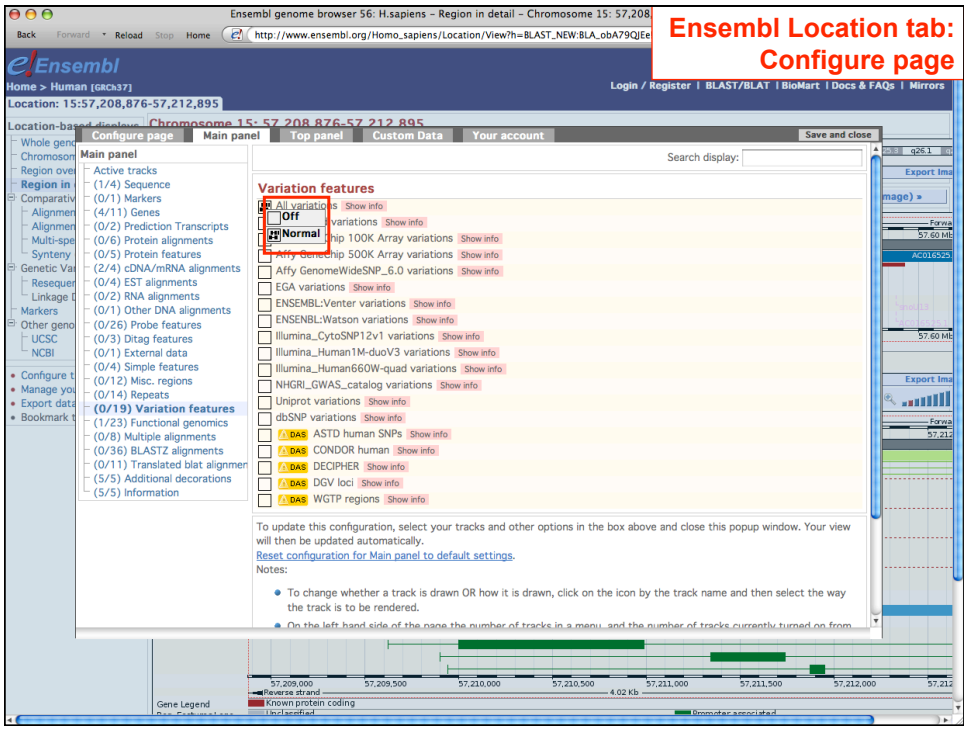

Ensembl BLAST search
BlastView
http://www.ensembl.org/Multi/blastview/BLA_obA79QJEe
Showing top 100 alignments of 2, sorted by Raw Score
refresh
Alignment Locations vs. Karyotype (click arrow to hide)
100% identity over 100% of the query length
Query Start 1
End 20
% ID 100.00
Alignment Locations vs. Query (click arrow to hide)
HSPs coverage
HSPs
Alignment Summary (click arrow to hide)
refresh
click
Subject
Chromosome
Supercontig
Clone
Contig
Chromosome
Chromosome
Chromosome
Stats
Sort By
Links
Query
Chromosome
Chromosome
Chromosome
Chromosome
Stats
[A] [S] [G] [C] 1 20 + Chr:15 57210876 57210895 + Chr:15 57210876 57210895 + Chr:15 57210876 57210895 + Chr:15 57210876 57210895 + 20 0.0093 100.00 20
[A] [S] [G] [C] 1 17 - Chr:8 72042559 72042575 + Chr:8 72042559 72042575 + Chr:8 72042559 72042575 + Chr:8 72042559 72042575 + 17 0.57 100.00 17
About Ensembl | Contact Us | Help

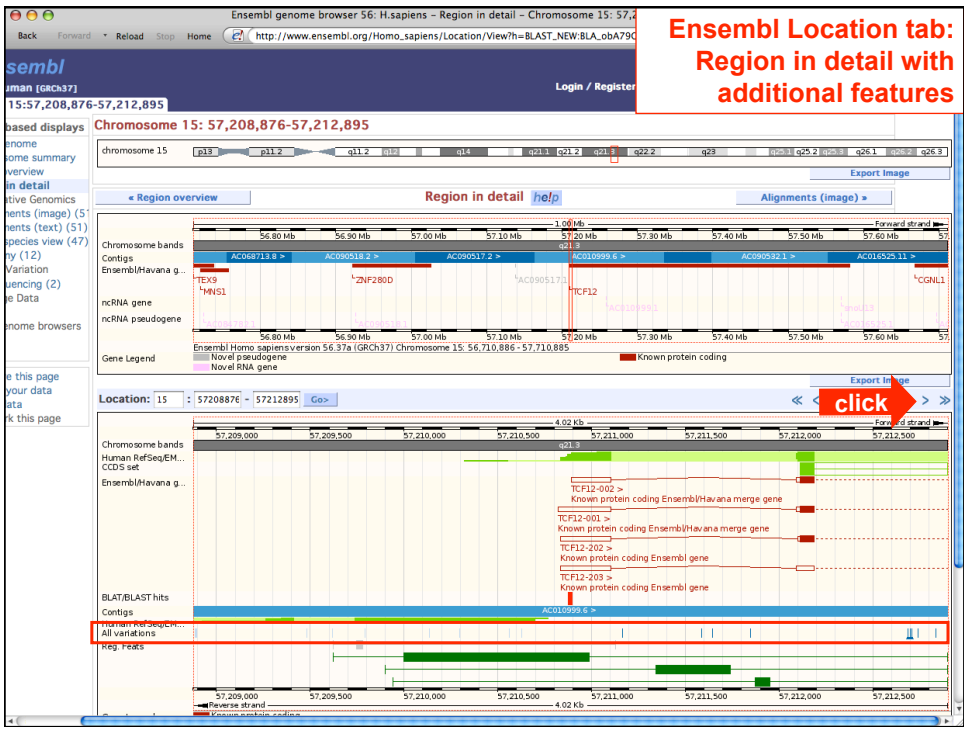


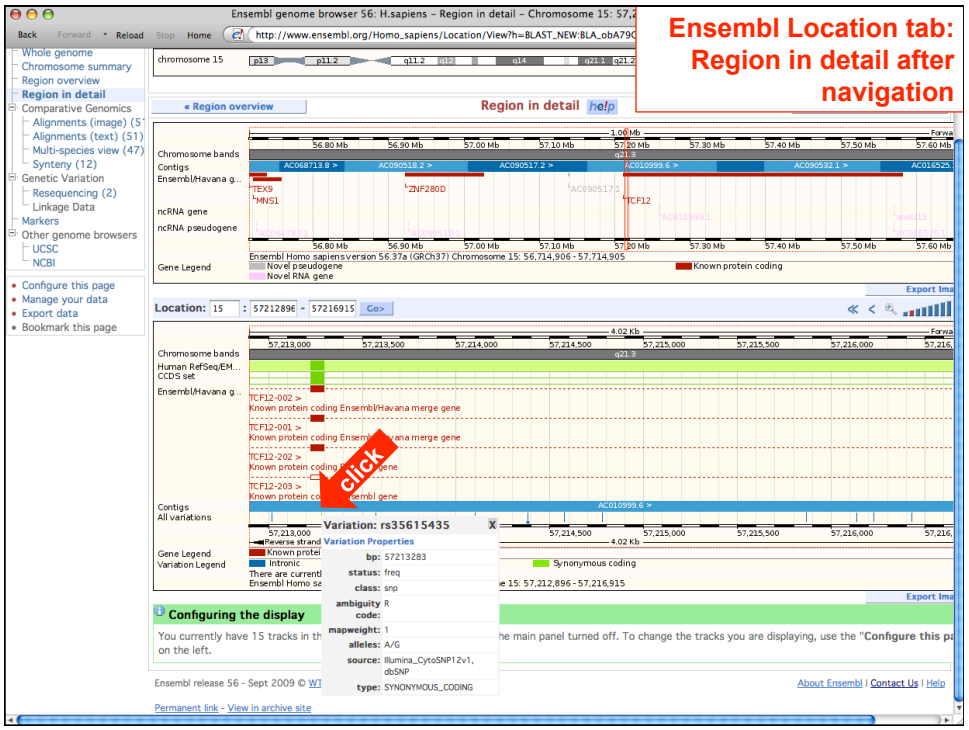
Ensembl

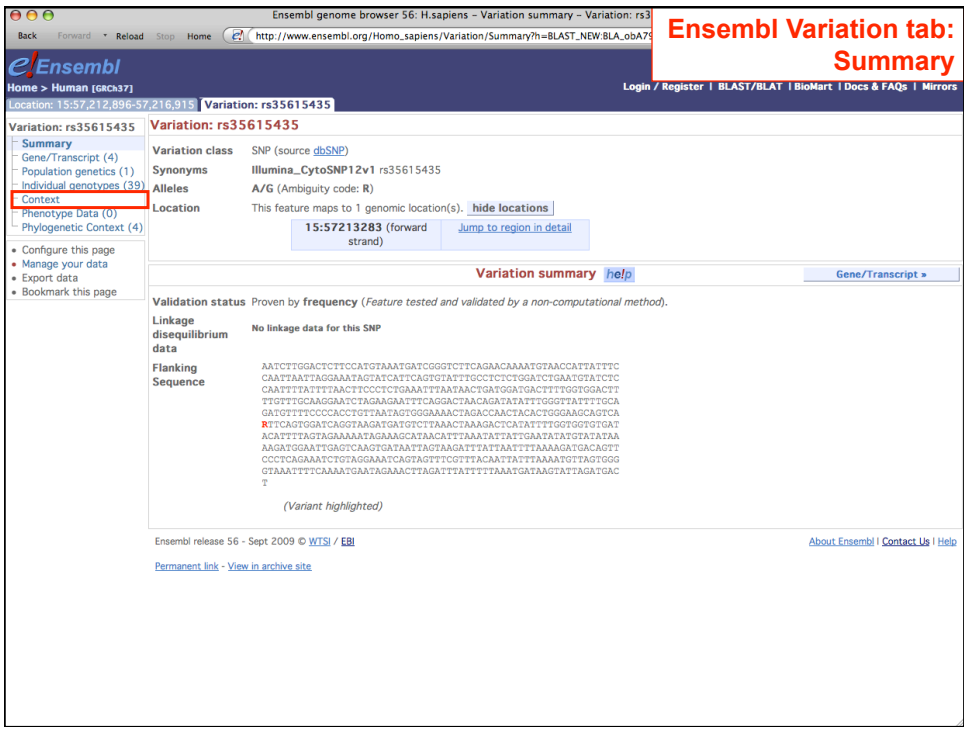
Location tab

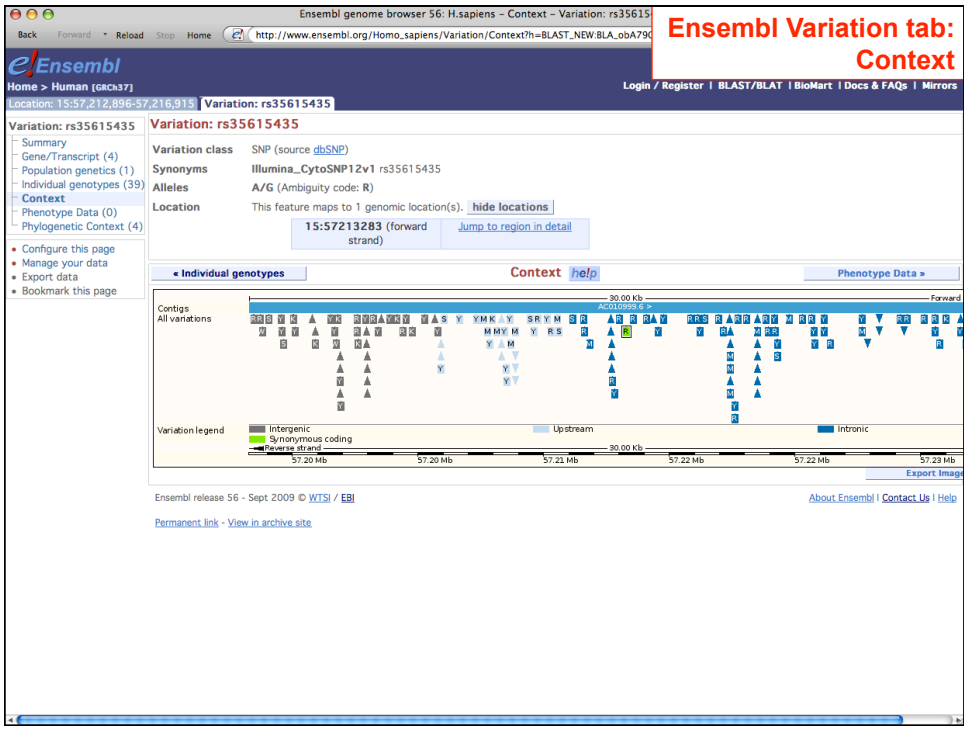

Ensembl Location tab:
Region in detail
Ensembl genome browser 56: H.sapiens – Region in detail – Chromosome 15: 57,208
Location: 15:57,208,876-57,212,895
Chromosome 15: 57,208,876-57,212,895
Region in detail
click
Configure this page
Manage your data
Export data
Bookmark this page


Ensembl Location tab:
Configure page
Ensembl genome browser 56: H.sapiens – Region in detail – Chromosome 15: 57,208
Location: 15:57,208,876-57,212,895
Configure page
Main panel
Top panel
Custom Data
Your account
Save and close
Variation features
Off
Normal
To update this configuration, select your tracks and other options in the box above and close this popup window. Your view will then be updated automatically.
Reset configuration for Main panel to default settings.
Notes:


Ensembl Location tab:
Region in detail with additional features
Chromosome 15: 57,208,876-57,212,895
Region in detail
click

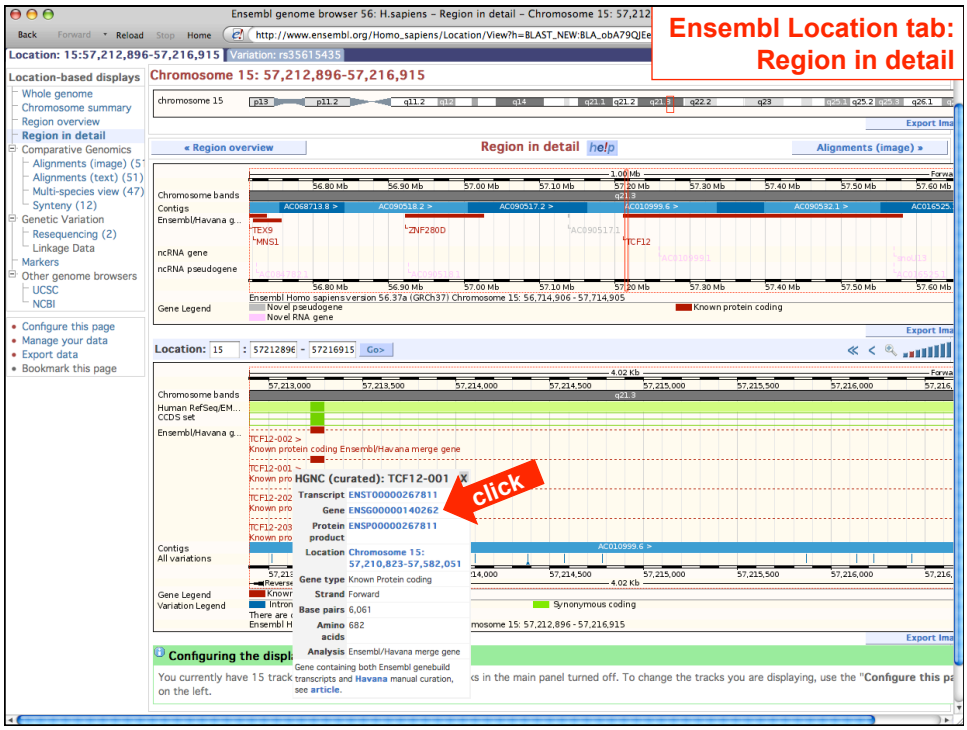

Ensembl Location tab:
Region in detail after navigation
Region in detail
click
Variation: rs35615435
Configuring the display

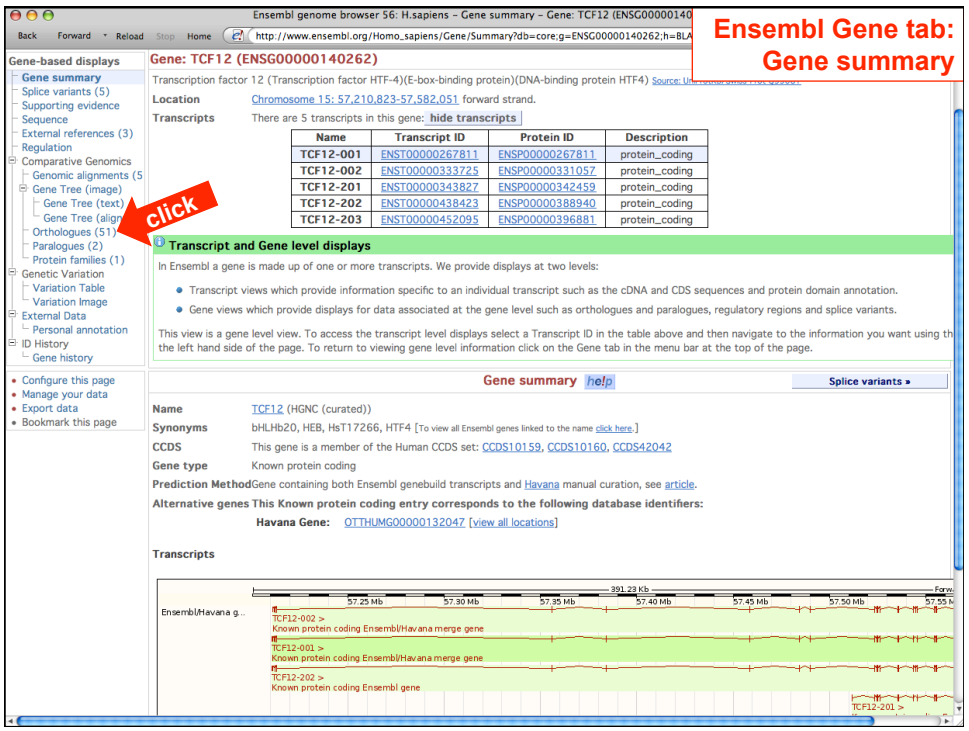

Ensembl Variation tab: Summary
Ensembl genome browser 56: H.sapiens – Variation summary – Variation: rs3
Variation: rs35615435
Summary
Gene/Transcript (4)
Population genetics (1)
Individual genotypes (39)
Context
Phenotype Data (0)
Phylogenetic Context (4)
Variation class SNP (source dbSNP)
Synonyms Illumina_CytoSNP12v1 rs35615435
Alleles A/G (Ambiguity code: R)
Location This feature maps to 1 genomic location(s). hide locations
15:57213283 (forward strand)
Jump to region in detail
Variation summary help
Validation status Proven by frequency (Feature tested and validated by a non-computational method).
Linkage disequilibrium data No linkage data for this SNP
Flanking Sequence
(Variant highlighted)

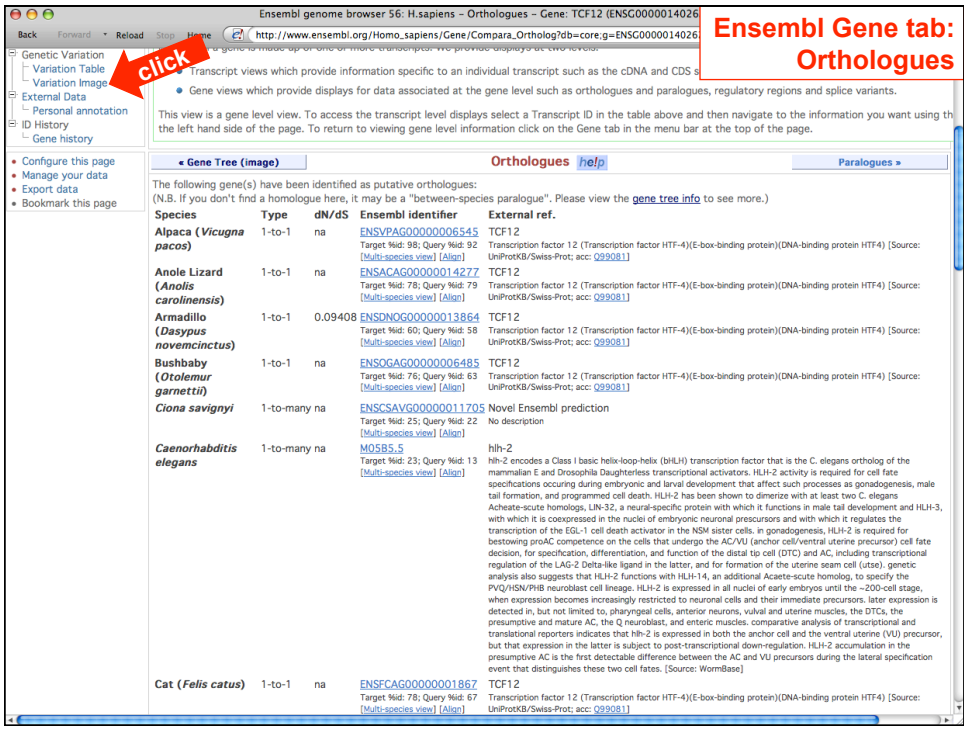

Ensembl Variation tab: Context
Ensembl genome browser 56: H.sapiens – Context – Variation: rs35615
Variation: rs35615435
Context
Variation class SNP (source dbSNP)
Synonyms Illumina_CytoSNP12v1 rs35615435
Alleles A/G (Ambiguity code: R)
Location This feature maps to 1 genomic location(s). hide locations
15:57213283 (forward strand)
Jump to region in detail
« Individual genotypes
Phenotype Data »
Contigs
All variations
Variation legend
Intergenic
Synonymous coding
Upstream
Intronic

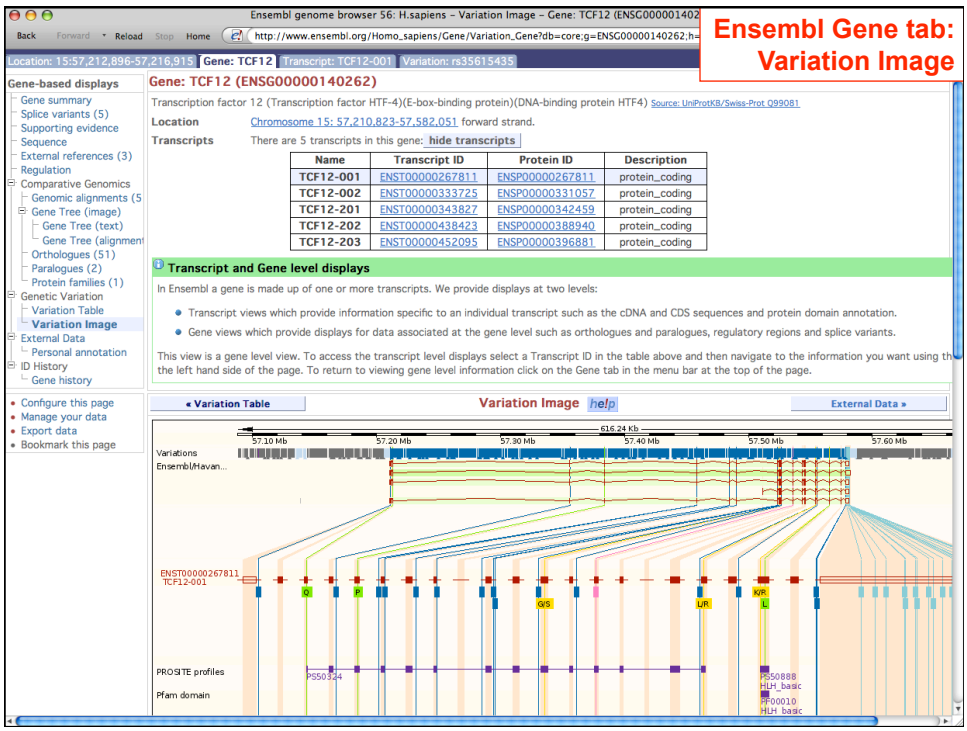
Ensembl

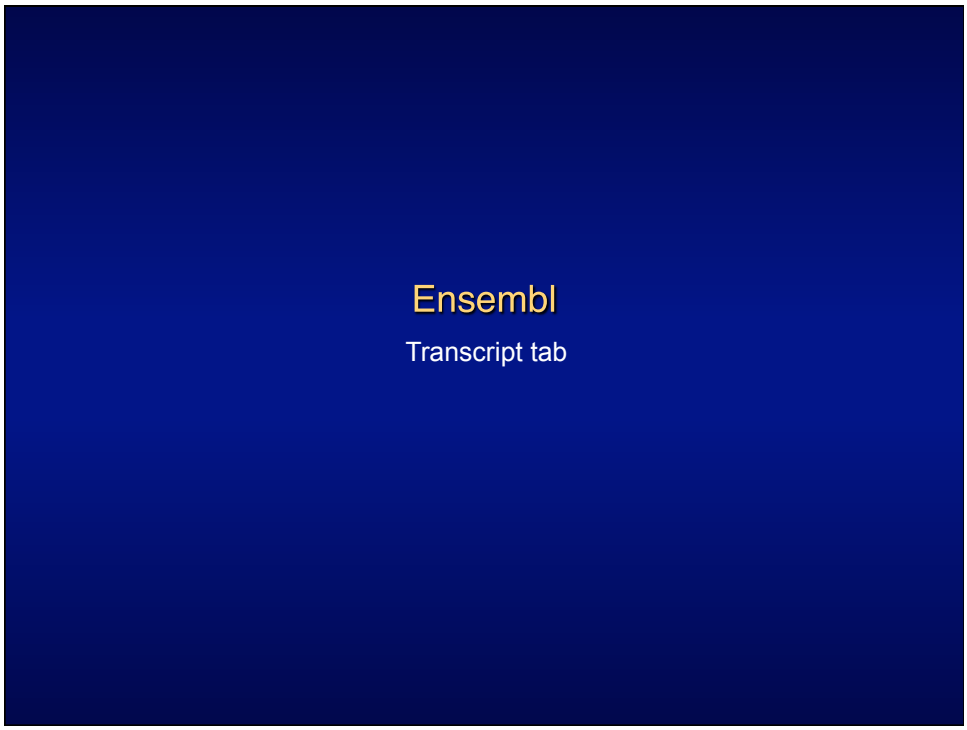
Gene tab

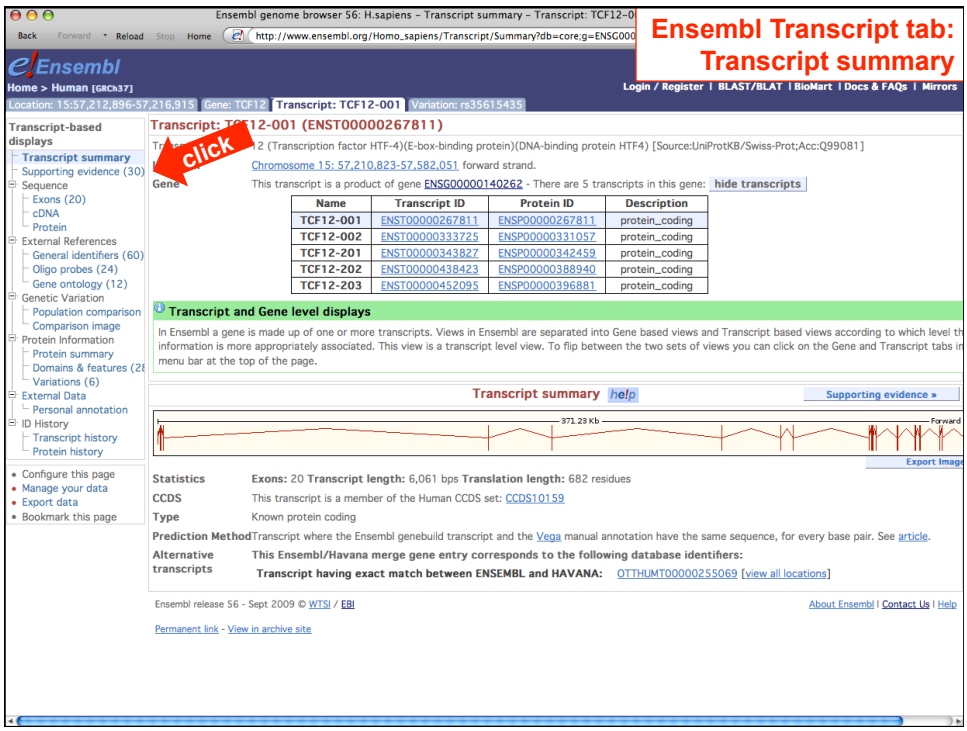


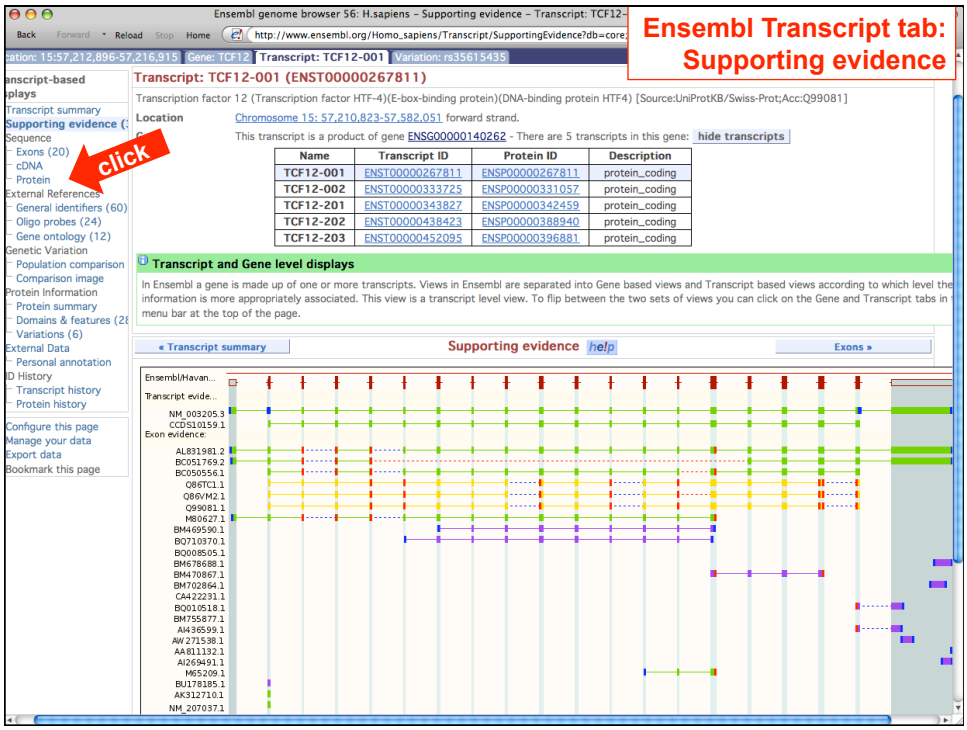

Ensembl Gene tab:
Gene summary
click
Gene: TCF12 (ENSG00000140262)
Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4)
Location
Chromosome 15: 57,210,823-57,582,051 forward strand.
Transcripts
There are 5 transcripts in this gene: hide transcripts
Name | Transcript ID | Protein ID | Description
TCF12-001 | ENST00000267811 | ENSP00000267811 | protein_coding
TCF12-002 | ENST00000333725 | ENSP00000331057 | protein_coding
TCF12-201 | ENST00000343827 | ENSP00000342459 | protein_coding
TCF12-202 | ENST00000438423 | ENSP00000388940 | protein_coding
TCF12-203 | ENST00000452095 | ENSP00000396881 | protein_coding
Transcript and Gene level displays
Gene summary help
Splice variants »
Name
TCF12 (HGNC (curated))
Synonyms
bHLHb20, HEB, HsT17266, HTF4
CCDS
This gene is a member of the Human CCDS set: CCDS10159, CCDS10160, CCDS42042
Gene type
Known protein coding
Prediction Method
Alternative genes
This Known protein coding entry corresponds to the following database identifiers:
Havana Gene:
OTTHUMG00000132047
Transcripts
TCF12-002 >
Known protein coding Ensembl/Havana merge gene
TCF12-001 >
Known protein coding Ensembl/Havana merge gene
TCF12-202 >
Known protein coding Ensembl gene
TCF12-201 >

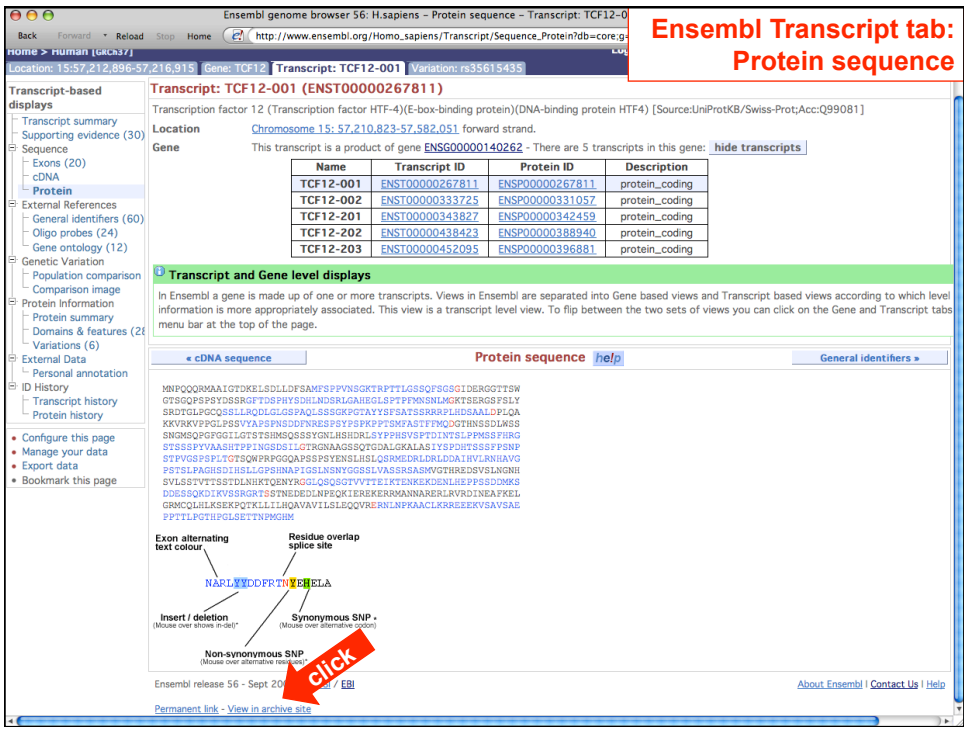

Ensembl Gene tab:
Orthologues
click
« Gene Tree (image)
Orthologues help
Paralogues »
The following gene(s) have been identified as putative orthologues:
(N.B. If you don't find a homologue here, it may be a "between-species paralogue". Please view the gene tree info to see more.)
Species | Type | dN/dS | Ensembl identifier | External ref.
Alpaca (Vicugna pacos) | 1-to-1 | na | ENSVPAG00000006545 | TCF12
Anole Lizard (Anolis carolinensis) | 1-to-1 | na | ENSACAG00000014277 | TCF12
Armadillo (Dasypus novemcinctus) | 1-to-1 | 0.09408 | ENSDNOG00000013864 | TCF12
Bushbaby (Otolemur garnettii) | 1-to-1 | na | ENSOGAG00000006485 | TCF12
Ciona savignyi | 1-to-many | na | ENSCSAVG00000011705 | Novel Ensembl prediction
Caenorhabditis elegans | 1-to-many | na | M05B5.5 | hlh-2
Cat (Felis catus) | 1-to-1 | na | ENSFCAG00000001867 | TCF12

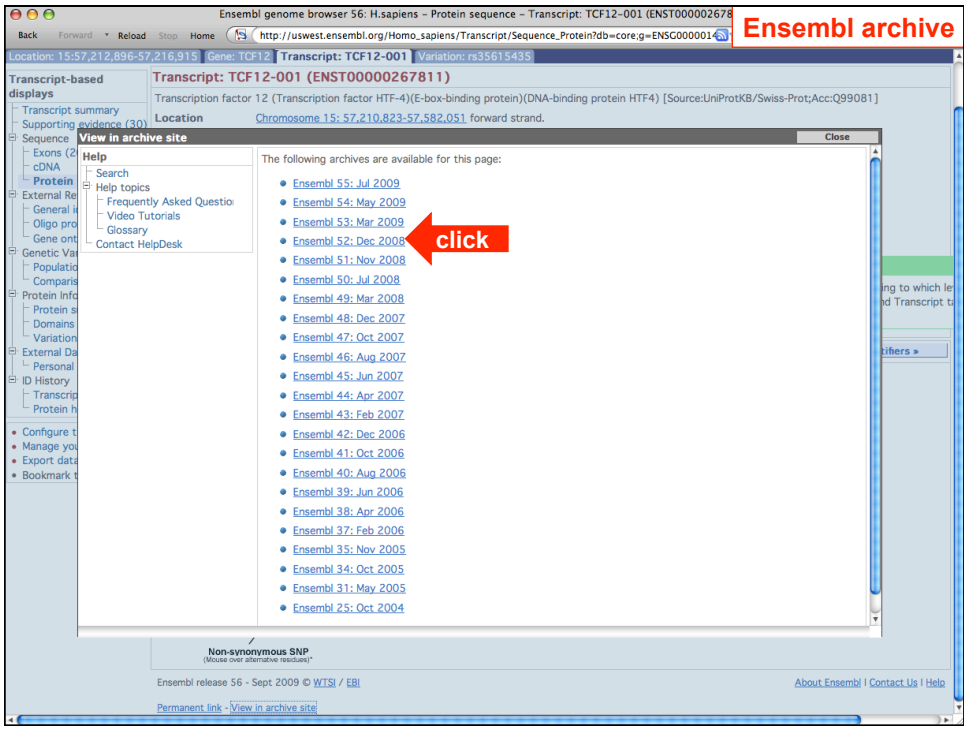

Ensembl Gene tab:
Variation Image
Ensembl genome browser 56: H.sapiens – Variation Image – Gene: TCF12 (ENSG00000140262
Gene: TCF12 (ENSG00000140262)
Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4)
Location
Chromosome 15: 57,210,823-57,582,051 forward strand.
Transcripts
There are 5 transcripts in this gene:
hide transcripts
Name
Transcript ID
Protein ID
Description
TCF12-001
ENST00000267811
ENSP00000267811
protein_coding
TCF12-002
ENST00000333725
ENSP00000331057
protein_coding
TCF12-201
ENST00000343827
ENSP00000342459
protein_coding
TCF12-202
ENST00000438423
ENSP00000388940
protein_coding
TCF12-203
ENST00000452095
ENSP00000396881
protein_coding
Transcript and Gene level displays
In Ensembl a gene is made up of one or more transcripts. We provide displays at two levels:
Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
Gene views which provide displays for data associated at the gene level such as orthologues and paralogues, regulatory regions and splice variants.
This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using th
the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.
Variation Image
help
Variations
PROSITE profiles
Pfam domain

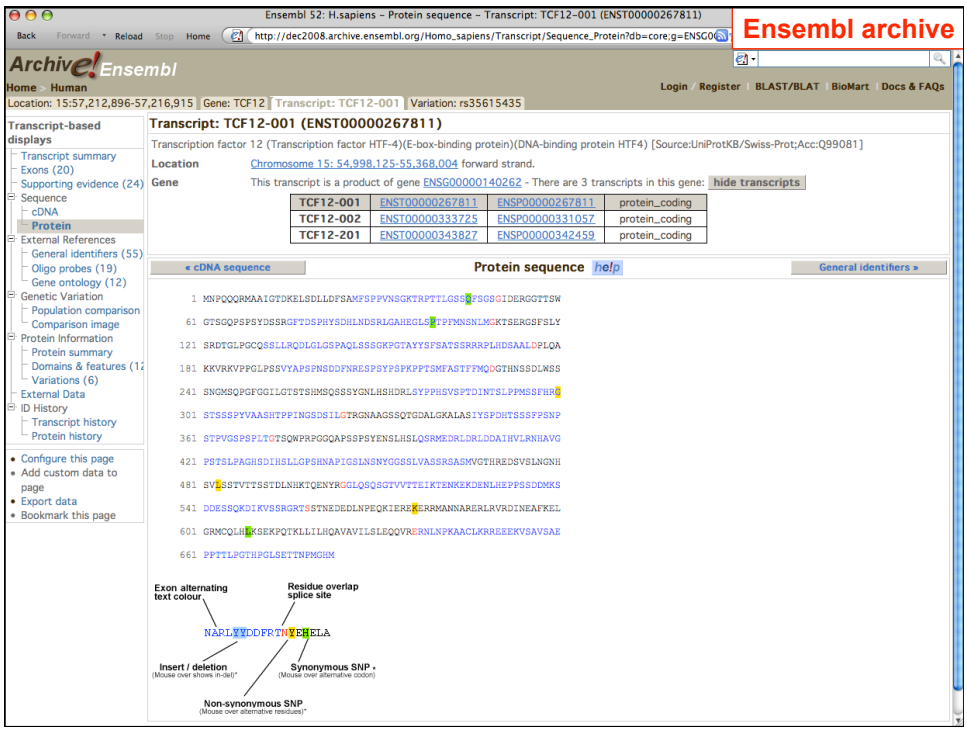

Ensembl
Transcript tab


Ensembl Transcript tab:
Transcript summary
click
Transcript: TCF12-001 (ENST00000267811)
Transcript summary
Transcript and Gene level displays


Ensembl Transcript tab:
Supporting evidence
click
Transcript: TCF12-001 (ENST00000267811)
Supporting evidence
Transcript and Gene level displays

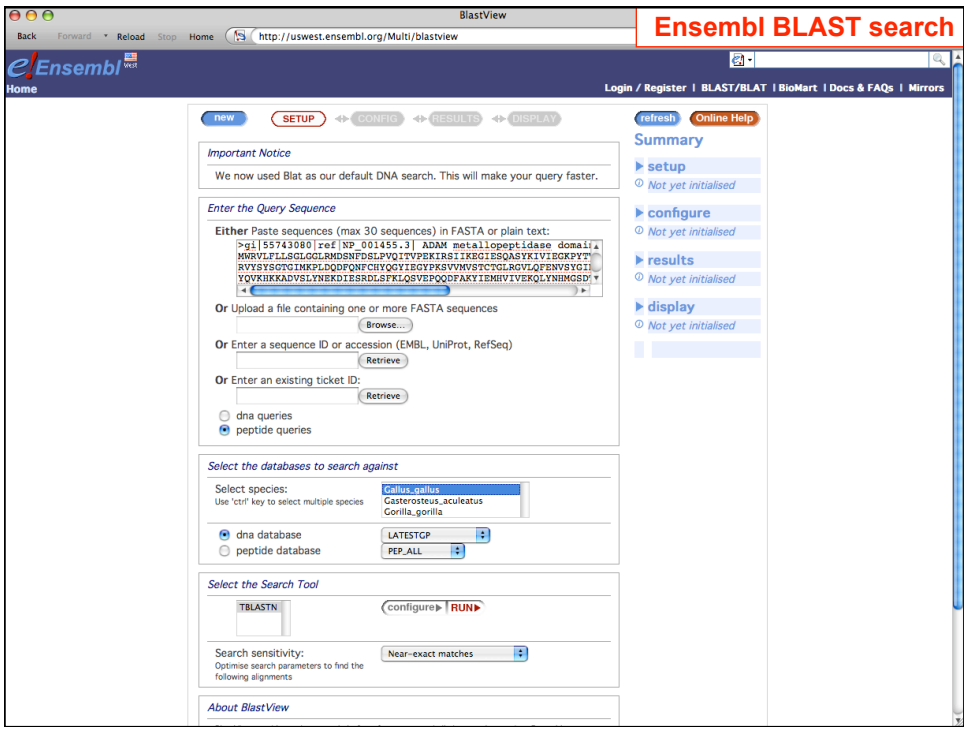

Ensembl Transcript tab:
Protein sequence
Ensembl genome browser 56: H.sapiens – Protein sequence – Transcript: TCF12-0
Transcript: TCF12-001 (ENST00000267811)
Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) [Source:UniProtKB/Swiss-Prot;Acc:Q99081]
Location Chromosome 15: 57,210,823-57,582,051 forward strand.
Gene This transcript is a product of gene ENSG00000140262 - There are 5 transcripts in this gene: hide transcripts
Name | Transcript ID | Protein ID | Description
TCF12-001 | ENST00000267811 | ENSP00000267811 | protein_coding
TCF12-002 | ENST00000333725 | ENSP00000331057 | protein_coding
TCF12-201 | ENST00000343827 | ENSP00000342459 | protein_coding
TCF12-202 | ENST00000438423 | ENSP00000388940 | protein_coding
TCF12-203 | ENST00000452095 | ENSP00000396881 | protein_coding
Transcript and Gene level displays
« cDNA sequence
Protein sequence help
General identifiers »
Exon alternating text colour
Residue overlap splice site
Insert / deletion
Synonymous SNP
Non-synonymous SNP
click
Permanent link - View in archive site
About Ensembl | Contact Us | Help

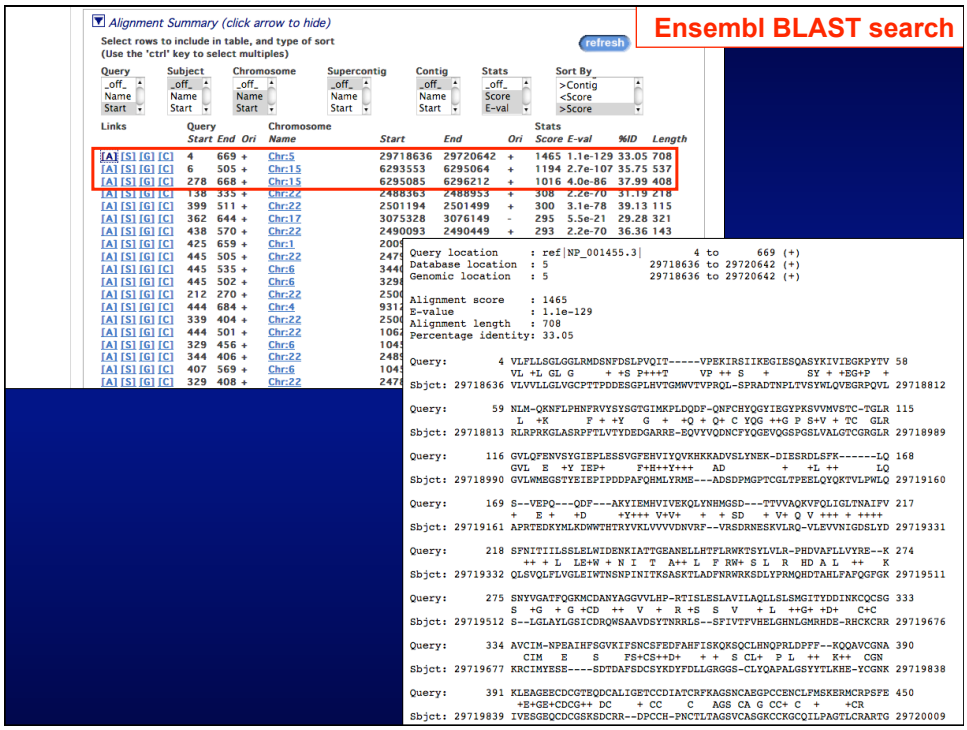

Ensembl archive
Ensembl genome browser 56: H.sapiens – Protein sequence – Transcript: TCF12-001 (ENST00000267811)
Transcript: TCF12-001 (ENST00000267811)
Transcription factor 12 (Transcription factor HTF-4)(E-box-binding protein)(DNA-binding protein HTF4) [Source:UniProtKB/Swiss-Prot;Acc:Q99081]
Location Chromosome 15: 57,210,823-57,582,051 forward strand.
View in archive site
Close
Help
Search
Help topics
Frequently Asked Questio
Video Tutorials
Glossary
Contact HelpDesk
The following archives are available for this page:
Ensembl 55: Jul 2009
Ensembl 54: May 2009
Ensembl 53: Mar 2009
Ensembl 52: Dec 2008
click
Ensembl 51: Nov 2008
Ensembl 50: Jul 2008
Ensembl 49: Mar 2008
Ensembl 48: Dec 2007
Ensembl 47: Oct 2007
Ensembl 46: Aug 2007
Ensembl 45: Jun 2007
Ensembl 44: Apr 2007
Ensembl 43: Feb 2007
Ensembl 42: Dec 2006
Ensembl 41: Oct 2006
Ensembl 40: Aug 2006
Ensembl 39: Jun 2006
Ensembl 38: Apr 2006
Ensembl 37: Feb 2006
Ensembl 35: Nov 2005
Ensembl 34: Oct 2005
Ensembl 31: May 2005
Ensembl 25: Oct 2004
Non-synonymous SNP
Ensembl release 56 - Sept 2009 © WTSI / EBI
About Ensembl | Contact Us | Help
Permanent link - View in archive site



Ensembl

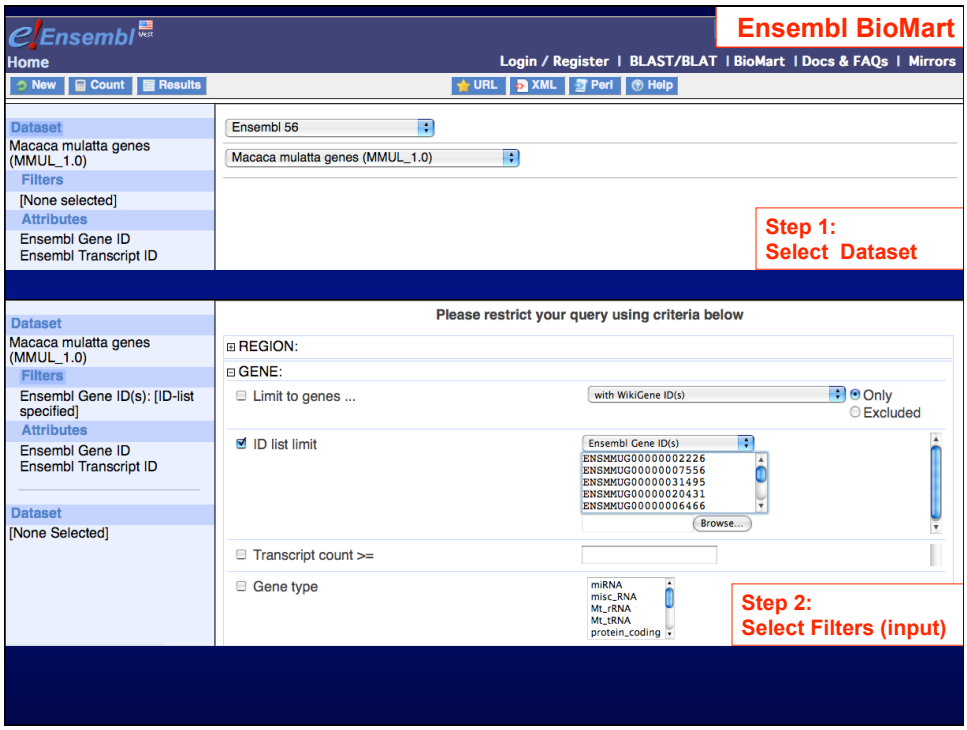
Find a chicken homolog of a human protein

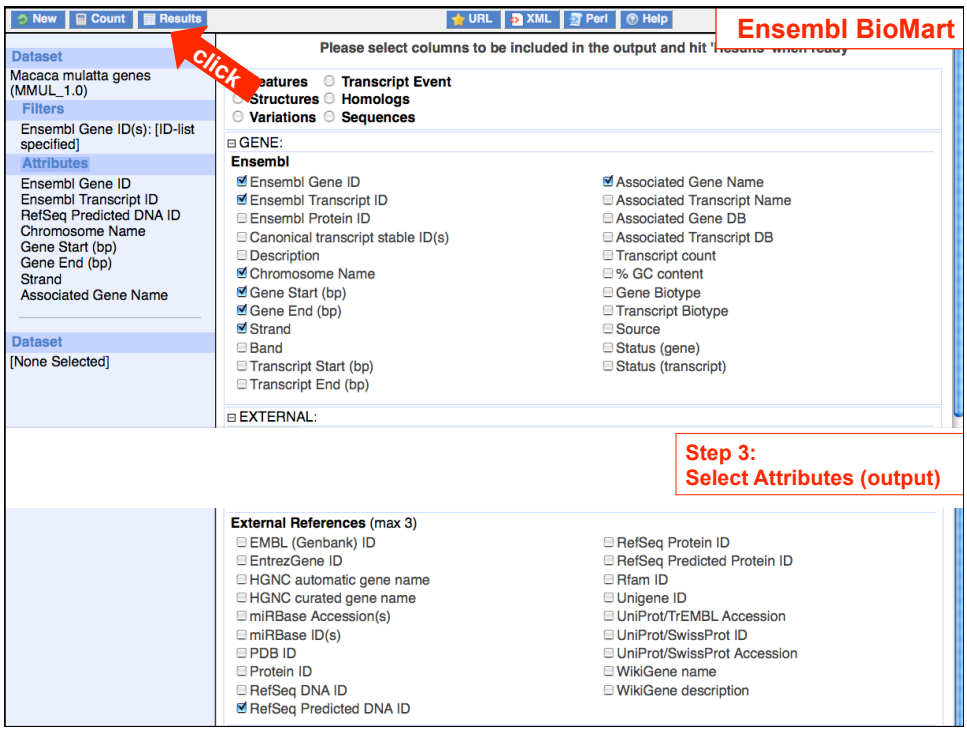

Ensembl BLAST search
BlastView
http://uswest.ensembl.org/Multi/blastview
Ensembl
Home
Login / Register | BLAST/BLAT | BioMart | Docs & FAQs | Mirrors
new
SETUP
CONFIG
RESULTS
DISPLAY
refresh
Online Help
Summary
setup
Not yet initialised
configure
Not yet initialised
results
Not yet initialised
display
Not yet initialised
Important Notice
We now used Blat as our default DNA search. This will make your query faster.
Enter the Query Sequence
Either Paste sequences (max 30 sequences) in FASTA or plain text:
>gi|55743080|ref|NP_001455.3| ADAM metallopeptidase domai
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKIVIEGKPYT
RVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGI
YQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDFAKYIEMHVIVEKQLYNHMGSD
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
Gorilla_gorilla
dna database
LATESTGP
peptide database
PEP_ALL
Select the Search Tool
TBLASTN
configure
RUN
Search sensitivity:
Optimise search parameters to find the following alignments
Near-exact matches
About BlastView

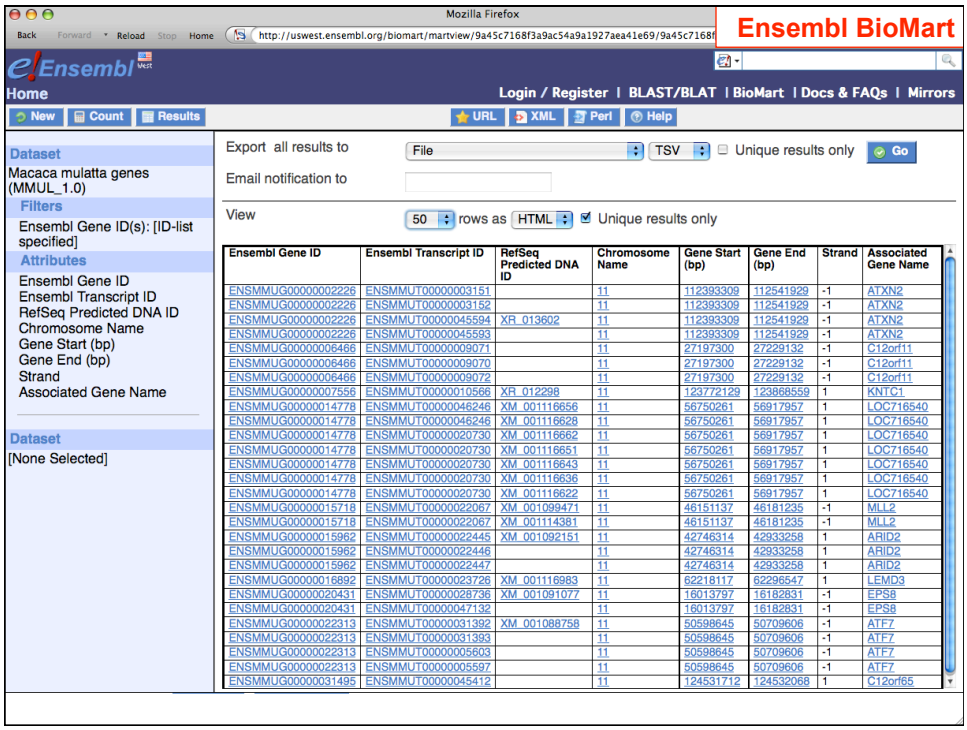

Ensembl BLAST search
Alignment Summary (click arrow to hide)
Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)
refresh
Query
Subject
Chromosome
Supercontig
Contig
Stats
Sort By
Links
Query Start End Ori
Chromosome Name
Start End Ori
Stats Score E-val %ID Length
[A] [S] [G] [C] 4 669 + Chr:5 29718636 29720642 + 1465 1.1e-129 33.05 708
[A] [S] [G] [C] 6 505 + Chr:15 6293553 6295064 + 1194 2.7e-107 35.75 537
[A] [S] [G] [C] 278 668 + Chr:15 6295085 6296212 + 1016 4.0e-86 37.99 408
[A] [S] [G] [C] 138 335 + Chr:22 2488363 2488953 + 308 2.2e-70 31.19 218
[A] [S] [G] [C] 399 511 + Chr:22 2501194 2501499 + 300 3.1e-78 39.13 115
[A] [S] [G] [C] 362 644 + Chr:17 3075328 3076149 - 295 5.5e-21 29.28 321
[A] [S] [G] [C] 438 570 + Chr:22 2490093 2490449 + 293 2.2e-70 36.36 143
[A] [S] [G] [C] 425 659 + Chr:1
[A] [S] [G] [C] 445 505 + Chr:22
[A] [S] [G] [C] 445 535 + Chr:6
[A] [S] [G] [C] 445 502 + Chr:6
[A] [S] [G] [C] 212 270 + Chr:22
[A] [S] [G] [C] 444 684 + Chr:4
[A] [S] [G] [C] 339 404 + Chr:22
[A] [S] [G] [C] 444 501 + Chr:22
[A] [S] [G] [C] 329 456 + Chr:6
[A] [S] [G] [C] 344 406 + Chr:22
[A] [S] [G] [C] 407 569 + Chr:6
[A] [S] [G] [C] 329 408 + Chr:22
Query location : ref|NP_001455.3| 4 to 669 (+)
Database location : 5 29718636 to 29720642 (+)
Genomic location : 5 29718636 to 29720642 (+)
Alignment score : 1465
E-value : 1.1e-129
Alignment length : 708
Percentage identity: 33.05
Query: 4 VLFLLSGLGGLRMDSNFDSLPVQIT-----VPEKIRSIIKEGIESQASYKIVIEGKPYTV 58
VL +L GL G + +S P+++T VP ++ S + SY + +EG+P +
Sbjct: 29718636 VLVVLLGLVGCPTTPDDESGPLHVTGMWVTVPRQL-SPRADTNPLTVSYWLQVEGRPQVL 29718812
Query: 59 NLM-QKNFLPHNFRVYSYSGTGIMKPLDQDF-QNFCHYQGYIEGYPKSVVMVSTC-TGLR 115
L +K F + +Y G + +Q + Q+ C YQG ++G P S+V + TC GLR
Sbjct: 29718813 RLRPRKGLASRPFTLVTYDEDGARRE-EQVYVQDNCFYQGEVQGSPGSLVALGTCGRGLR 29718989
Query: 116 GVLQFENVSYGIEPLESSVGFEHVIYQVKHKKADVSLYNEK-DIESRDLSFK------LQ 168
GVL E +Y IEP+ F+H++Y+++ AD + +L ++ LQ
Sbjct: 29718990 GVLWMEGSTYEIEPIPDDPAFQHMLYRME---ADSDPMGPTCGLTPEELQYQKTVLPWLQ 29719160
Query: 169 S--VEPQ---QDF---AKYIEMHVIVEKQLYNHMGSD---TTVVAQKVFQLIGLTNAIFV 217
+ E + +D +Y+++ V+V+ + + SD + V+ Q V +++ + ++++
Sbjct: 29719161 APRTEDKYMLKDWWTHTRYVKLVVVVDNVRF--VRSDRNESKVLRQ-VLEVVNIGDSLYD 29719331
Query: 218 SFNITTIILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLR-PHDVAFLLVYRE--K 274
++ + L LE+W + N I T A++ L F RW+ S L R HD A L ++ K
Sbjct: 29719332 QLSVQLFLVGLEIWTNSNPINITKSASKTLADFNRWRKSDLYPRMQHDTAHLFAFQGFGK 29719511
Query: 275 SNYVGATFQGKMCDANYAGGVVLHP-RTISLESLAVILAQLLSLSMGITYDDINKCQCSG 333
S +G + G +CD ++ V + R +S S V + L ++G+ +D+ C+C
Sbjct: 29719512 S--LGLAYLGSICDRQWSAAVDSYTNRRLS--SFIVTFVHELGHNLGMRHDE-RHCKCRR 29719676
Query: 334 AVCIM-NPEAIHFSGVKIFSNCSFEDFAHFISKQKSQCLHNQPRLDPFF--KQQAVCGNA 390
CIM E S FS+CS++D+ + + S CL+ P L ++ K++ CGN
Sbjct: 29719677 KRCIMYESE----SDTDAFSDCSYKDYFDLLGRGGS-CLYQAPALGSYYTLKHE-YCGNK 29719838
Query: 391 KLEAGEECDCGTEQDCALIGETCCDIATCRFKAGSNCAEGPCCENCLFMSKERMCRPSFE 450
+E+GE+CDCG++ DC + CC C AGS CA G CC+ C + +CR
Sbjct: 29719839 IVESGEQCDCGSKSDCRR--DPCCH-PNCTLTAGSVCASGKCCKGCQILPAGTLCRARTG 29720009

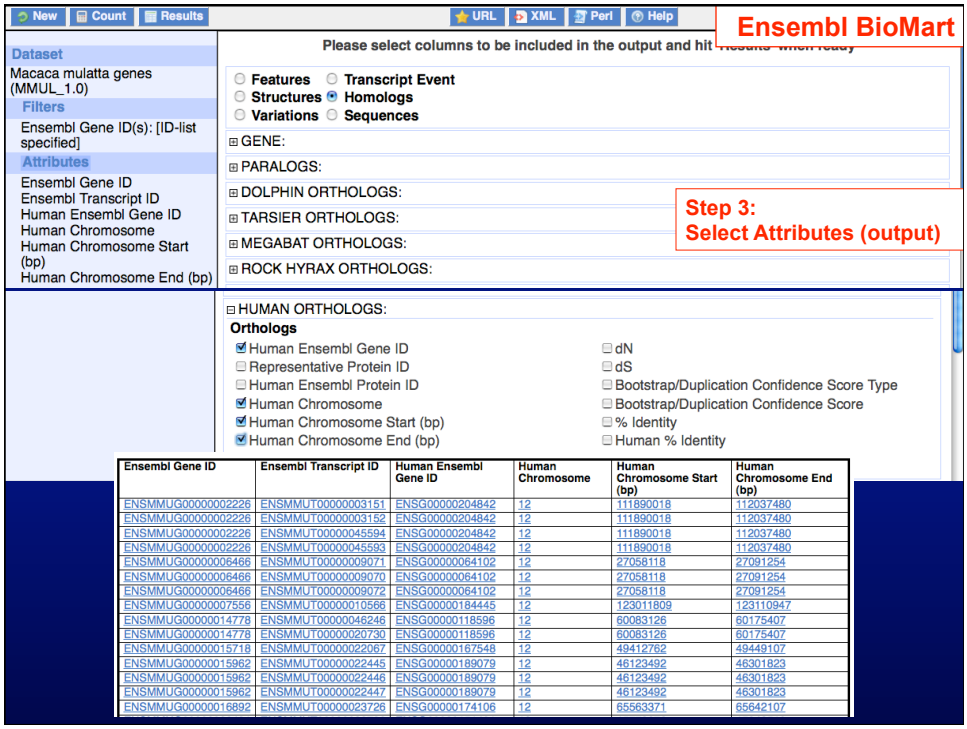
Ensembl

Using BioMart to cross-reference data from different sources




New
Count
Results
URL
XML
Perl
Help
Ensembl BioMart
click
Dataset
Macaca mulatta genes (MMUL_1.0)
Filters
Ensembl Gene ID(s): [ID-list specified]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
RefSeq Predicted DNA ID
Chromosome Name
Gene Start (bp)
Gene End (bp)
Strand
Associated Gene Name
Dataset
[None Selected]
Please select columns to be included in the output and hit
Transcript Event
Homologs
Sequences
GENE:
Ensembl
Ensembl Gene ID
Ensembl Transcript ID
Ensembl Protein ID
Canonical transcript stable ID(s)
Description
Chromosome Name
Gene Start (bp)
Gene End (bp)
Strand
Band
Transcript Start (bp)
Transcript End (bp)
Associated Gene Name
Associated Transcript Name
Associated Gene DB
Associated Transcript DB
Transcript count
% GC content
Gene Biotype
Transcript Biotype
Source
Status (gene)
Status (transcript)
EXTERNAL:
Step 3:
Select Attributes (output)
External References (max 3)
EMBL (Genbank) ID
EntrezGene ID
HGNC automatic gene name
HGNC curated gene name
miRBase Accession(s)
miRBase ID(s)
PDB ID
Protein ID
RefSeq DNA ID
RefSeq Predicted DNA ID
RefSeq Protein ID
RefSeq Predicted Protein ID
Rfam ID
Unigene ID
UniProt/TrEMBL Accession
UniProt/SwissProt ID
UniProt/SwissProt Accession
WikiGene name
WikiGene description

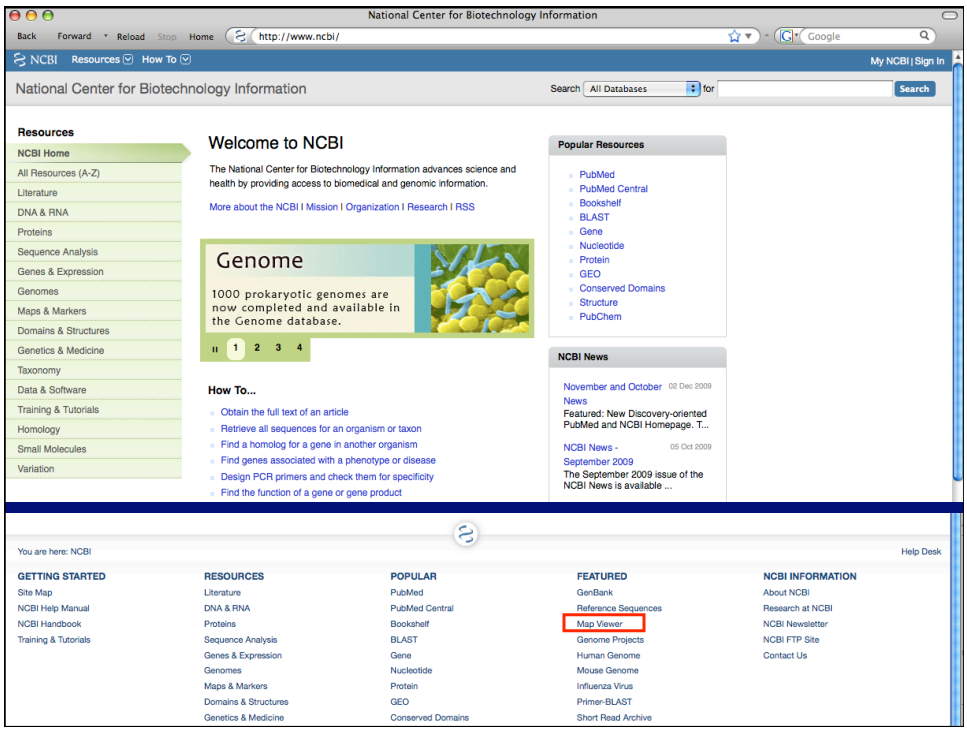

Mozilla Firefox
Ensembl BioMart
Ensembl
Home
Login / Register | BLAST/BLAT | BioMart | Docs & FAQs | Mirrors
New
Count
Results
URL
XML
Perl
Help
Dataset
Macaca mulatta genes (MMUL_1.0)
Filters
Ensembl Gene ID(s): [ID-list specified]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
RefSeq Predicted DNA ID
Chromosome Name
Gene Start (bp)
Gene End (bp)
Strand
Associated Gene Name
Dataset
[None Selected]
Export all results to
File
TSV
Unique results only
Go
Email notification to
View
50
rows as
HTML
Unique results only
Ensembl Gene ID
Ensembl Transcript ID
RefSeq Predicted DNA ID
Chromosome Name
Gene Start (bp)
Gene End (bp)
Strand
Associated Gene Name
ENSMMUG00000002226 ENSMMUT00000003151 11 112393309 112541929 -1 ATXN2
ENSMMUG00000002226 ENSMMUT00000003152 11 112393309 112541929 -1 ATXN2
ENSMMUG00000002226 ENSMMUT00000045594 XR_013602 11 112393309 112541929 -1 ATXN2
ENSMMUG00000002226 ENSMMUT00000045593 11 112393309 112541929 -1 ATXN2
ENSMMUG00000006466 ENSMMUT00000009071 11 27197300 27229132 -1 C12orf11
ENSMMUG00000006466 ENSMMUT00000009070 11 27197300 27229132 -1 C12orf11
ENSMMUG00000006466 ENSMMUT00000009072 11 27197300 27229132 -1 C12orf11
ENSMMUG00000007556 ENSMMUT00000010566 XR_012298 11 123772129 123868559 1 KNTC1
ENSMMUG00000014778 ENSMMUT00000046246 XM_001116656 11 56750261 56917957 1 LOC716540
ENSMMUG00000014778 ENSMMUT00000046246 XM_001116628 11 56750261 56917957 1 LOC716540
ENSMMUG00000014778 ENSMMUT00000020730 XM_001116662 11 56750261 56917957 1 LOC716540
ENSMMUG00000014778 ENSMMUT00000020730 XM_001116651 11 56750261 56917957 1 LOC716540
ENSMMUG00000014778 ENSMMUT00000020730 XM_001116643 11 56750261 56917957 1 LOC716540
ENSMMUG00000014778 ENSMMUT00000020730 XM_001116636 11 56750261 56917957 1 LOC716540
ENSMMUG00000014778 ENSMMUT00000020730 XM_001116622 11 56750261 56917957 1 LOC716540
ENSMMUG00000015718 ENSMMUT00000022067 XM_001099471 11 46151137 46181235 -1 MLL2
ENSMMUG00000015718 ENSMMUT00000022067 XM_001114381 11 46151137 46181235 -1 MLL2
ENSMMUG00000015962 ENSMMUT00000022445 XM_001092151 11 42746314 42933258 1 ARID2
ENSMMUG00000015962 ENSMMUT00000022446 11 42746314 42933258 1 ARID2
ENSMMUG00000015962 ENSMMUT00000022447 11 42746314 42933258 1 ARID2
ENSMMUG00000016892 ENSMMUT00000023726 XM_001116983 11 62218117 62296547 1 LEMD3
ENSMMUG00000020431 ENSMMUT00000028736 XM_001091077 11 16013797 16182831 -1 EPS8
ENSMMUG00000020431 ENSMMUT00000047132 11 16013797 16182831 -1 EPS8
ENSMMUG00000022313 ENSMMUT00000031392 XM_001088758 11 50598645 50709606 -1 ATF7
ENSMMUG00000022313 ENSMMUT00000031393 11 50598645 50709606 -1 ATF7
ENSMMUG00000022313 ENSMMUT00000005603 11 50598645 50709606 -1 ATF7
ENSMMUG00000022313 ENSMMUT00000005597 11 50598645 50709606 -1 ATF7
ENSMMUG00000031495 ENSMMUT00000045412 11 124531712 124532068 1 C12orf65

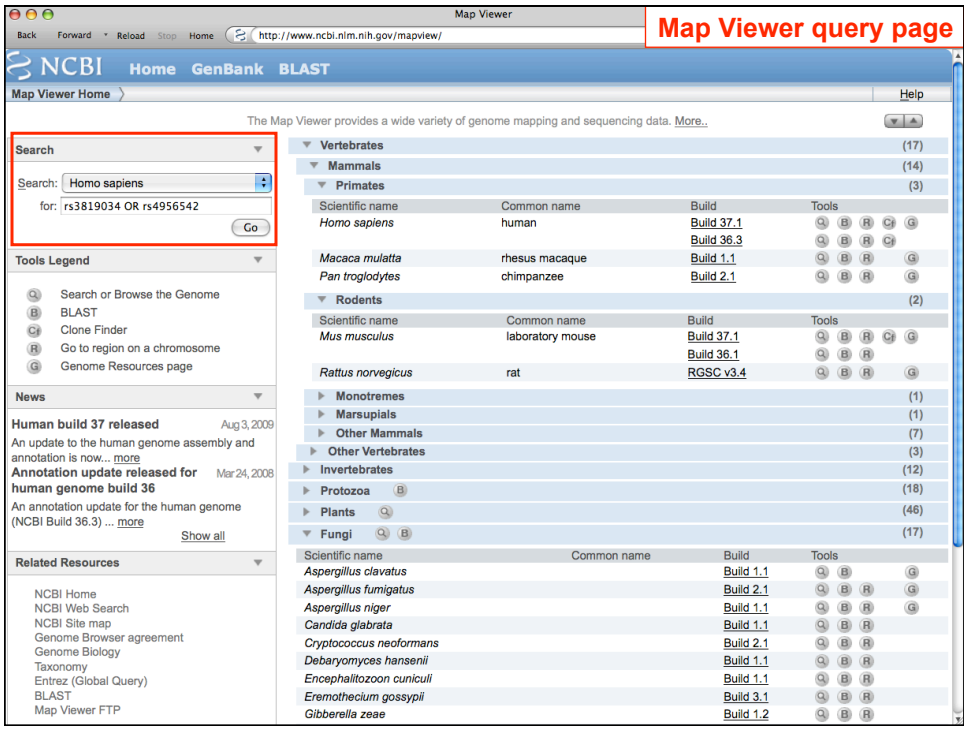


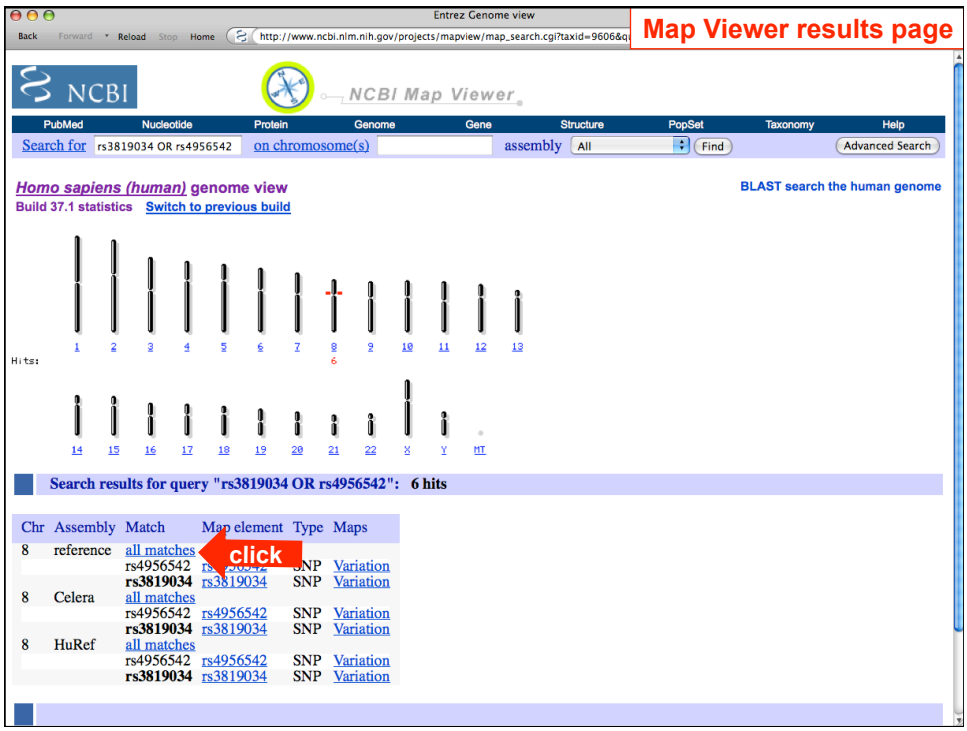
Ensembl Gene ID	Ensembl Transcript ID	Human Ensembl Gene ID	Human Chromosome	Human Chromosome Start (bp)	Human Chromosome End (bp)
ENSMMUG00000002226	ENSMMUT00000003151	ENSG00000204842	12	111890018	112037480
ENSMMUG00000002226	ENSMMUT00000003152	ENSG00000204842	12	111890018	112037480
ENSMMUG00000002226	ENSMMUT00000045594	ENSG00000204842	12	111890018	112037480
ENSMMUG00000002226	ENSMMUT00000045593	ENSG00000204842	12	111890018	112037480
ENSMMUG00000006466	ENSMMUT00000009071	ENSG00000064102	12	27058118	27091254
ENSMMUG00000006466	ENSMMUT00000009070	ENSG00000064102	12	27058118	27091254
ENSMMUG00000006466	ENSMMUT00000009072	ENSG00000064102	12	27058118	27091254
ENSMMUG00000007556	ENSMMUT00000010566	ENSG00000184445	12	123011809	123110947
ENSMMUG00000014778	ENSMMUT00000046246	ENSG00000118596	12	60083126	60175407
ENSMMUG00000014778	ENSMMUT00000020730	ENSG00000118596	12	60083126	60175407
ENSMMUG00000015718	ENSMMUT00000022067	ENSG00000167548	12	49412762	49449107
ENSMMUG00000015962	ENSMMUT00000022445	ENSG00000189079	12	46123492	46301823
ENSMMUG00000015962	ENSMMUT00000022446	ENSG00000189079	12	46123492	46301823
ENSMMUG00000015962	ENSMMUT00000022447	ENSG00000189079	12	46123492	46301823
ENSMMUG00000016892	ENSMMUT00000023726	ENSG00000174106	12	65563371	65642107

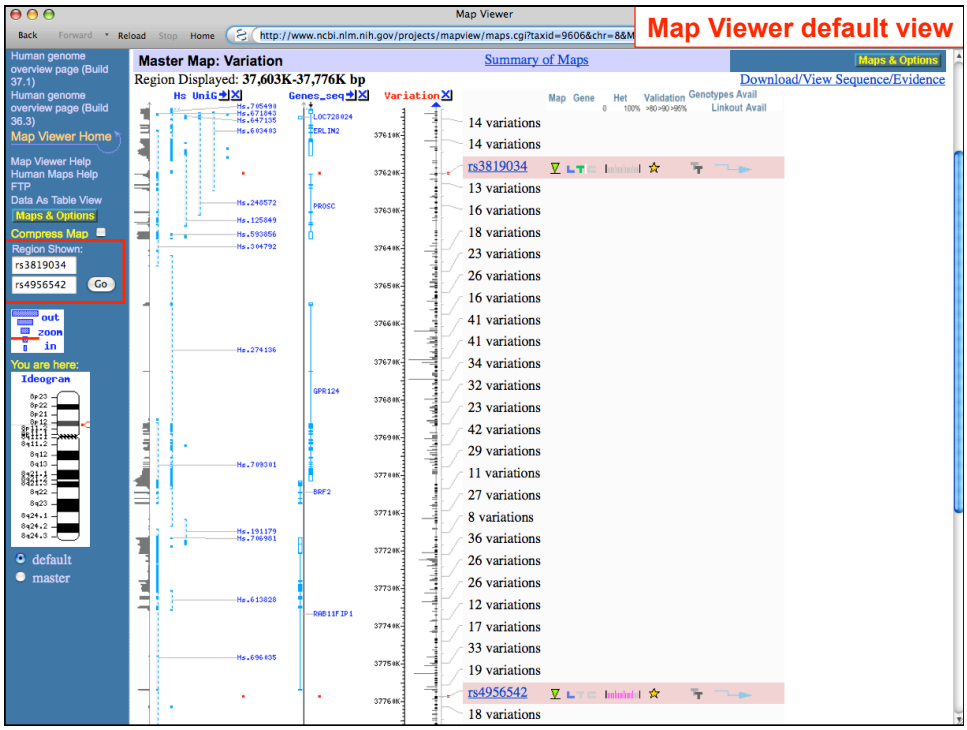
NCBI

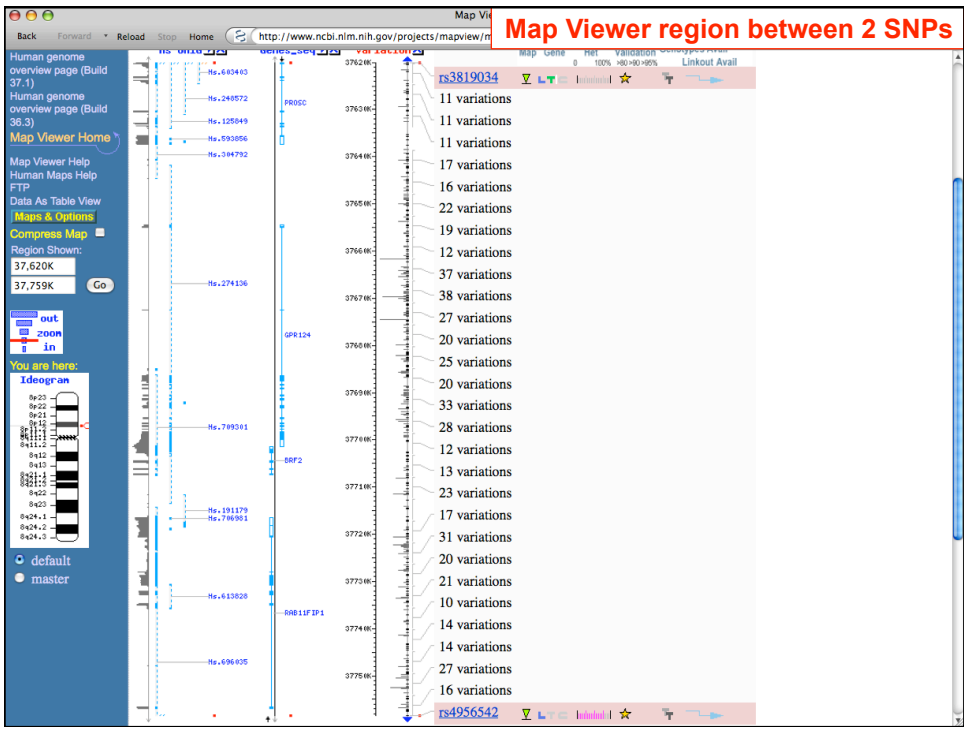
View a genomic region between two SNPs


National Center for Biotechnology Information
http://www.ncbi/
NCBI Resources How To
My NCBI | Sign In
National Center for Biotechnology Information
Search All Databases for Search
Resources
NCBI Home
All Resources (A-Z)
Literature
DNA & RNA
Proteins
Sequence Analysis
Genes & Expression
Genomes
Maps & Markers
Domains & Structures
Genetics & Medicine
Taxonomy
Data & Software
Training & Tutorials
Homology
Small Molecules
Variation
Welcome to NCBI
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
More about the NCBI | Mission | Organization | Research | RSS
Genome
1000 prokaryotic genomes are now completed and available in the Genome database.
How To...
Obtain the full text of an article
Retrieve all sequences for an organism or taxon
Find a homolog for a gene in another organism
Find genes associated with a phenotype or disease
Design PCR primers and check them for specificity
Find the function of a gene or gene product
Popular Resources
PubMed
PubMed Central
Bookshelf
BLAST
Gene
Nucleotide
Protein
GEO
Conserved Domains
Structure
PubChem
NCBI News
November and October News
Featured: New Discovery-oriented PubMed and NCBI Homepage. T...
NCBI News - September 2009
The September 2009 issue of the NCBI News is available ...
You are here: NCBI
Help Desk
GETTING STARTED
Site Map
NCBI Help Manual
NCBI Handbook
Training & Tutorials
RESOURCES
Literature
DNA & RNA
Proteins
Sequence Analysis
Genes & Expression
Genomes
Maps & Markers
Domains & Structures
Genetics & Medicine
POPULAR
PubMed
PubMed Central
Bookshelf
BLAST
Gene
Nucleotide
Protein
GEO
Conserved Domains
FEATURED
GenBank
Reference Sequences
Map Viewer
Genome Projects
Human Genome
Mouse Genome
Influenza Virus
Primer-BLAST
Short Read Archive
NCBI INFORMATION
About NCBI
Research at NCBI
NCBI Newsletter
NCBI FTP Site
Contact Us


Map Viewer
http://www.ncbi.nlm.nih.gov/mapview/
Map Viewer query page
NCBI Home GenBank BLAST
Map Viewer Home
Help
The Map Viewer provides a wide variety of genome mapping and sequencing data. More..
Search
Search: Homo sapiens
for: rs3819034 OR rs4956542
Go
Tools Legend
Search or Browse the Genome
BLAST
Clone Finder
Go to region on a chromosome
Genome Resources page
News
Human build 37 released Aug 3, 2009
An update to the human genome assembly and annotation is now... more
Annotation update released for human genome build 36 Mar 24, 2008
An annotation update for the human genome (NCBI Build 36.3) ... more
Show all
Related Resources
NCBI Home
NCBI Web Search
NCBI Site map
Genome Browser agreement
Genome Biology
Taxonomy
Entrez (Global Query)
BLAST
Map Viewer FTP
Vertebrates (17)
Mammals (14)
Primates (3)
Scientific name | Common name | Build | Tools
Homo sapiens | human | Build 37.1, Build 36.3
Macaca mulatta | rhesus macaque | Build 1.1
Pan troglodytes | chimpanzee | Build 2.1
Rodents (2)
Mus musculus | laboratory mouse | Build 37.1, Build 36.1
Rattus norvegicus | rat | RGSC v3.4
Monotremes (1)
Marsupials (1)
Other Mammals (7)
Other Vertebrates (3)
Invertebrates (12)
Protozoa (18)
Plants (46)
Fungi (17)
Scientific name | Common name | Build | Tools
Aspergillus clavatus | Build 1.1
Aspergillus fumigatus | Build 2.1
Aspergillus niger | Build 1.1
Candida glabrata | Build 1.1
Cryptococcus neoformans | Build 2.1
Debaryomyces hansenii | Build 1.1
Encephalitozoon cuniculi | Build 1.1
Eremothecium gossypii | Build 3.1
Gibberella zeae | Build 1.2

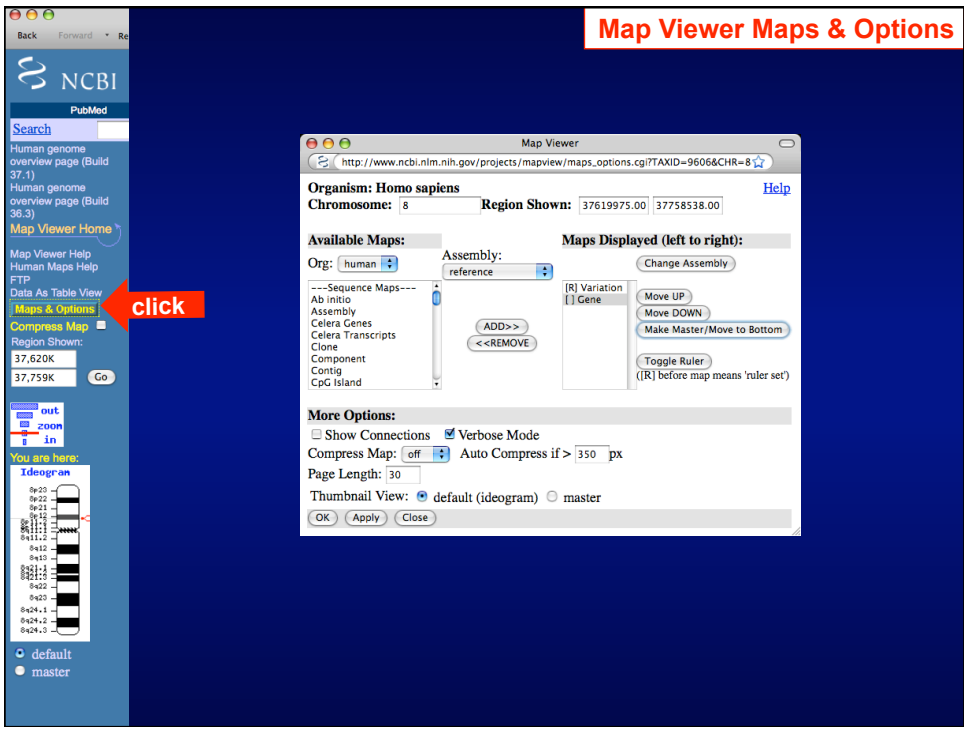

Map Viewer results page
Entrez Genome view
NCBI
NCBI Map Viewer
PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help
Search for rs3819034 OR rs4956542
on chromosome(s)
assembly All
Find
Advanced Search
Homo sapiens (human) genome view
Build 37.1 statistics
Switch to previous build
BLAST search the human genome
Hits:
Search results for query "rs3819034 OR rs4956542": 6 hits
Chr Assembly Match Map element Type Maps
8 reference all matches
rs4956542 rs4956542 SNP Variation
rs3819034 rs3819034 SNP Variation
8 Celera all matches
rs4956542 rs4956542 SNP Variation
rs3819034 rs3819034 SNP Variation
8 HuRef all matches
rs4956542 rs4956542 SNP Variation
rs3819034 rs3819034 SNP Variation
click

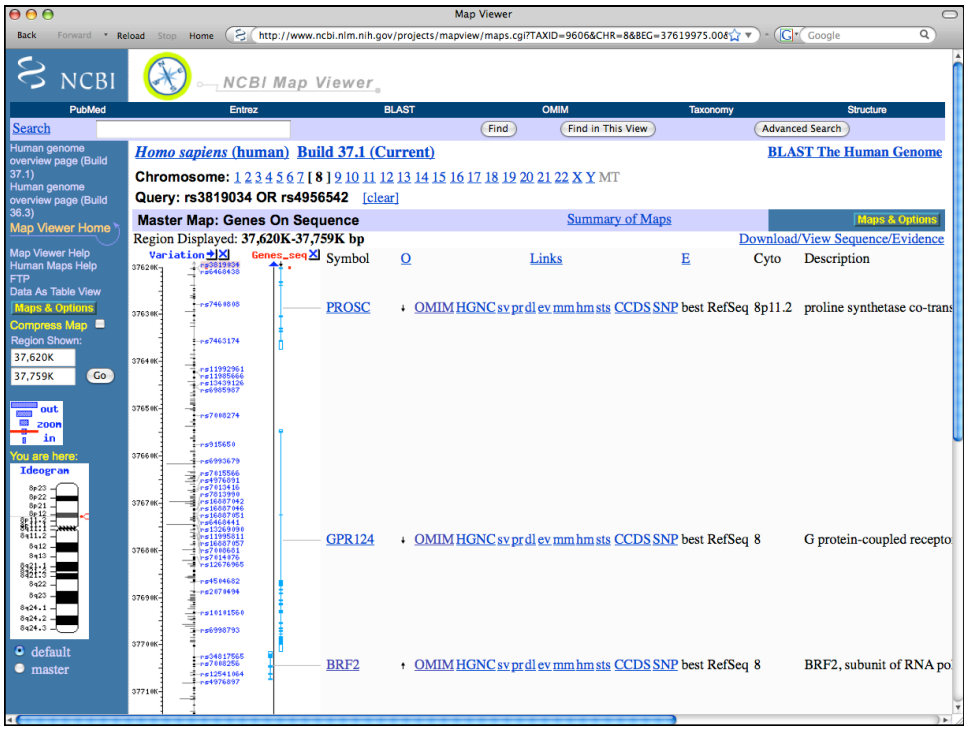

Map Viewer default view
Map Viewer
Master Map: Variation
Summary of Maps
Maps & Options
Region Displayed: 37,603K-37,776K bp
Download/View Sequence/Evidence
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Compress Map
Region Shown:
rs3819034
rs4956542
Go
You are here:
Ideogram
default
master
14 variations
14 variations
rs3819034
13 variations
16 variations
18 variations
23 variations
26 variations
16 variations
41 variations
41 variations
34 variations
32 variations
23 variations
42 variations
29 variations
11 variations
27 variations
8 variations
36 variations
26 variations
26 variations
12 variations
17 variations
33 variations
19 variations
rs4956542
18 variations



NCBI

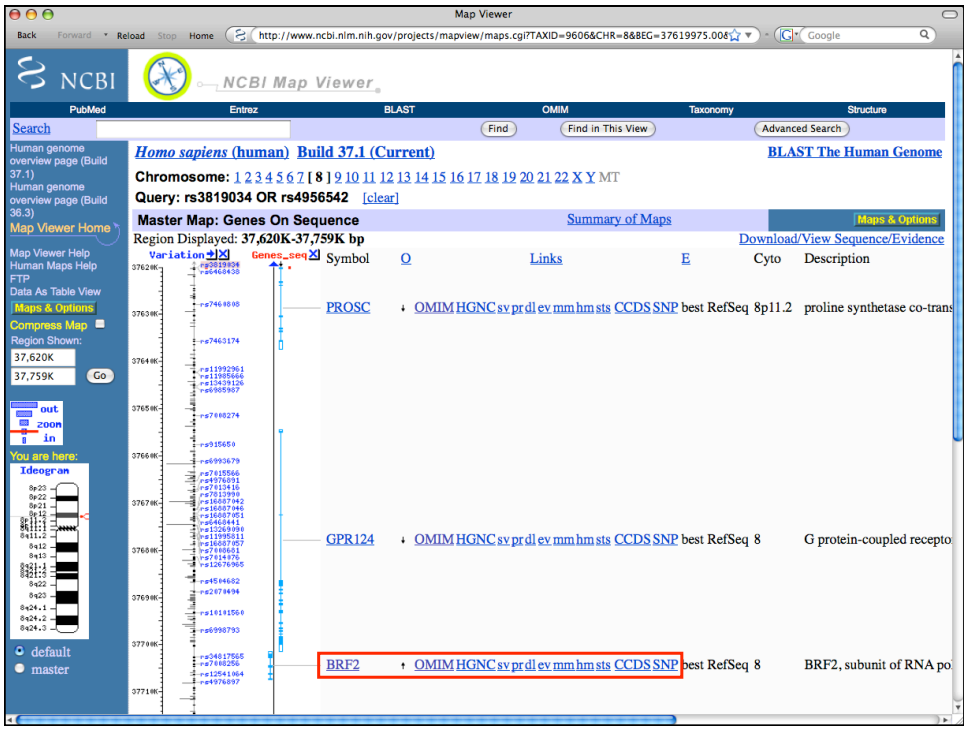
Change the maps displayed on the Map Viewer

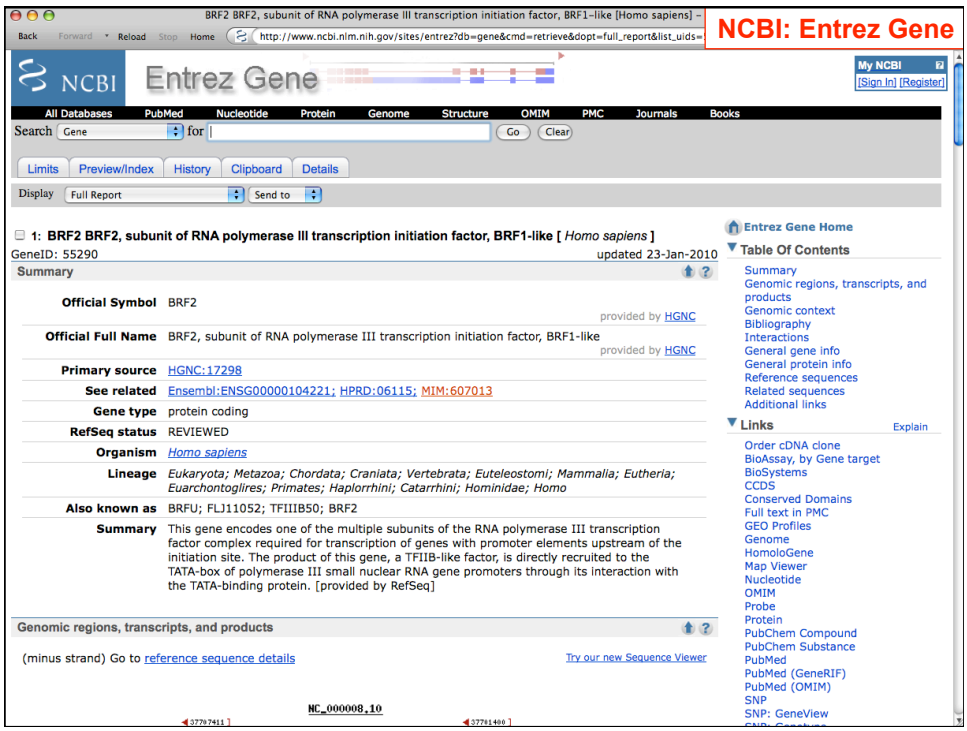

Map Viewer Maps & Options
NCBI
PubMed
Search
Human genome overview page (Build 37.1)
Human genome overview page (Build 36.3)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
37,620K
37,759K
Go
out
zoom
in
You are here:
Ideogram
default
master
click
Map Viewer
http://www.ncbi.nlm.nih.gov/projects/mapview/maps_options.cgi?TAXID=9606&CHR=8
Organism: Homo sapiens
Help
Chromosome: 8
Region Shown: 37619975.00 37758538.00
Available Maps:
Org: human
Assembly:
reference
---Sequence Maps---
Ab initio Assembly
Celera Genes
Celera Transcripts
Clone Component
Contig
CpG Island
ADD>>
<<REMOVE
Maps Displayed (left to right):
Change Assembly
[R] Variation
[ ] Gene
Move UP
Move DOWN
Make Master/Move to Bottom
Toggle Ruler
([R] before map means 'ruler set')
More Options:
Show Connections
Verbose Mode
Compress Map: off
Auto Compress if > 350 px
Page Length: 30
Thumbnail View: default (ideogram) master
OK
Apply
Close

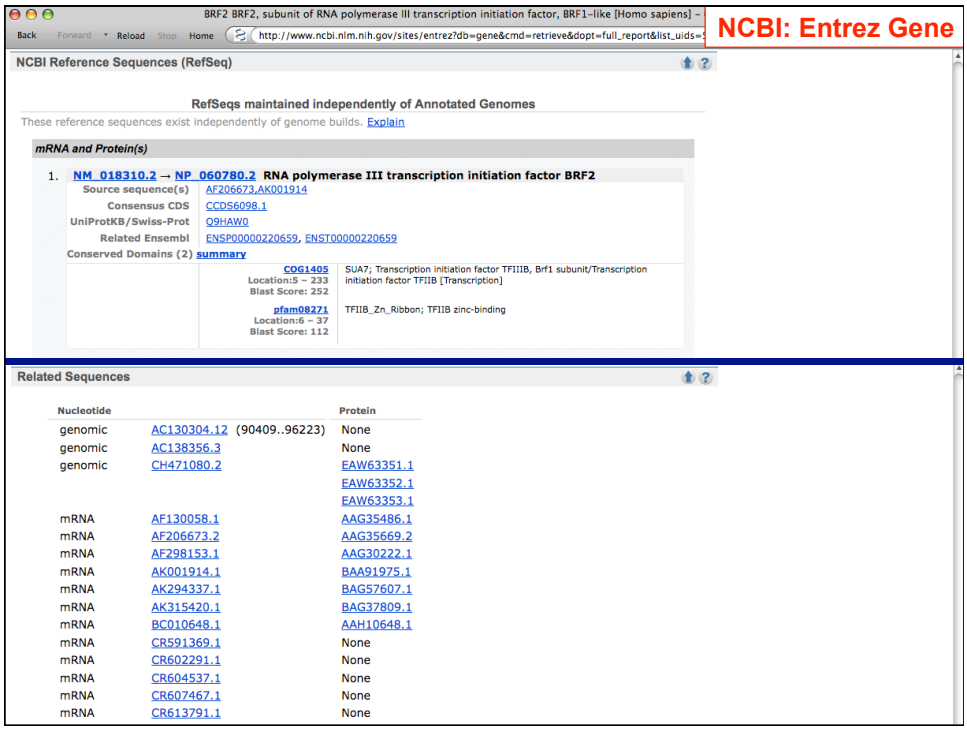

Map Viewer
NCBI
NCBI Map Viewer
PubMed
Entrez
BLAST
OMIM
Taxonomy
Structure
Search
Find
Find in This View
Advanced Search
Homo sapiens (human) Build 37.1 (Current)
BLAST The Human Genome
Chromosome: 1 2 3 4 5 6 7 [8] 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: rs3819034 OR rs4956542 [clear]
Master Map: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 37,620K-37,759K bp
Download/View Sequence/Evidence
Variation
Genes_seq
Symbol
Links
Cyto
Description
PROSC
GPR124
BRF2
OMIM HGNC sv pr dl ev mm hm sts CCDS SNP best RefSeq
8p11.2
proline synthetase co-trans
G protein-coupled recepto
BRF2, subunit of RNA po

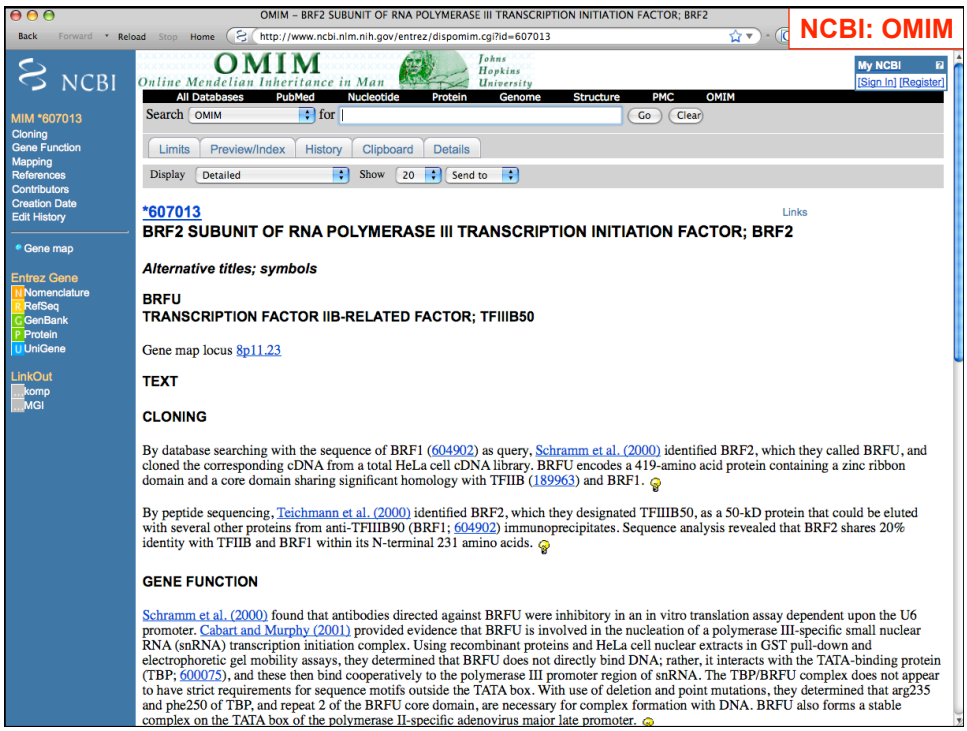
NCBI

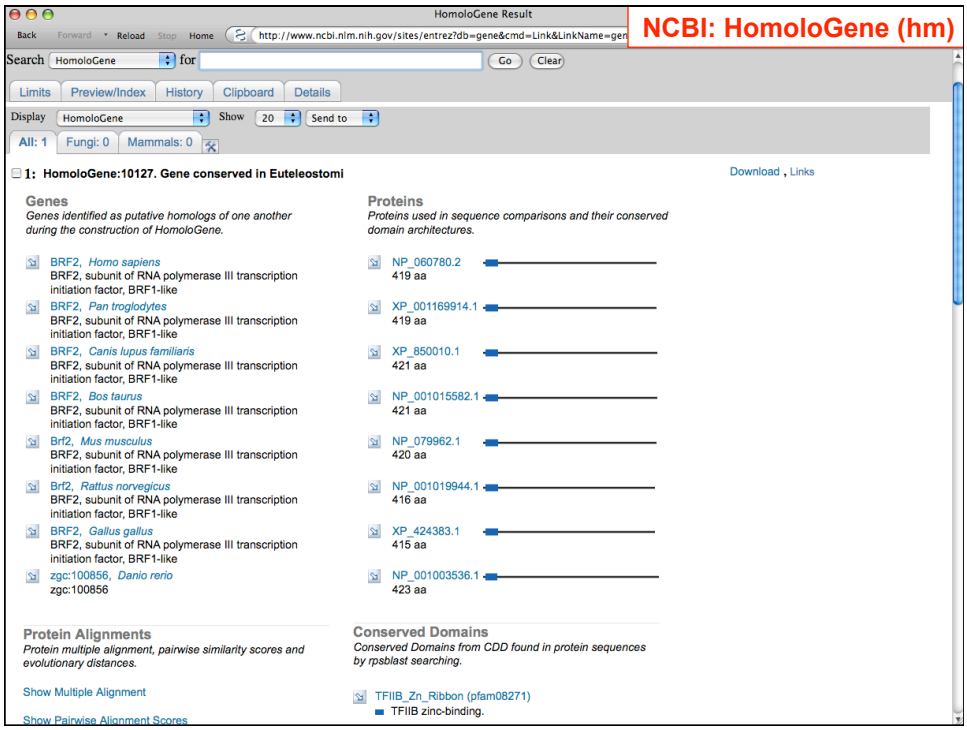
View additional information about a gene




NCBI: Entrez Gene
BRF2 BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like [Homo sapiens] -
Entrez Gene
1: BRF2 BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like [Homo sapiens]
GeneID: 55290
updated 23-Jan-2010
Summary
Official Symbol BRF2
Official Full Name BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
Primary source HGNC:17298
See related Ensembl:ENSG00000104221; HPRD:06115; MIM:607013
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 BRFU; FLJ11052; TFIIIB50; BRF2
Summary This gene encodes one of the multiple subunits of the RNA polymerase III transcription factor complex required for transcription of genes with promoter elements upstream of the initiation site. The product of this gene, a TFIIB-like factor, is directly recruited to the TATA-box of polymerase III small nuclear RNA gene promoters through its interaction with the TATA-binding protein. [provided by RefSeq]
Genomic regions, transcripts, and products
(minus strand) Go to reference sequence details
Try our new Sequence Viewer
NC_000008.10
Entrez Gene Home
Table Of Contents
Summary
Genomic regions, transcripts, and products
Genomic context
Bibliography
Interactions
General gene info
General protein info
Reference sequences
Related sequences
Additional links
Links
Order cDNA clone
BioAssay, by Gene target
BioSystems
CCDS
Conserved Domains
Full text in PMC
GEO Profiles
Genome
HomoloGene
Map Viewer
Nucleotide
OMIM
Probe
Protein
PubChem Compound
PubChem Substance
PubMed
PubMed (GeneRIF)
PubMed (OMIM)
SNP
SNP: GeneView

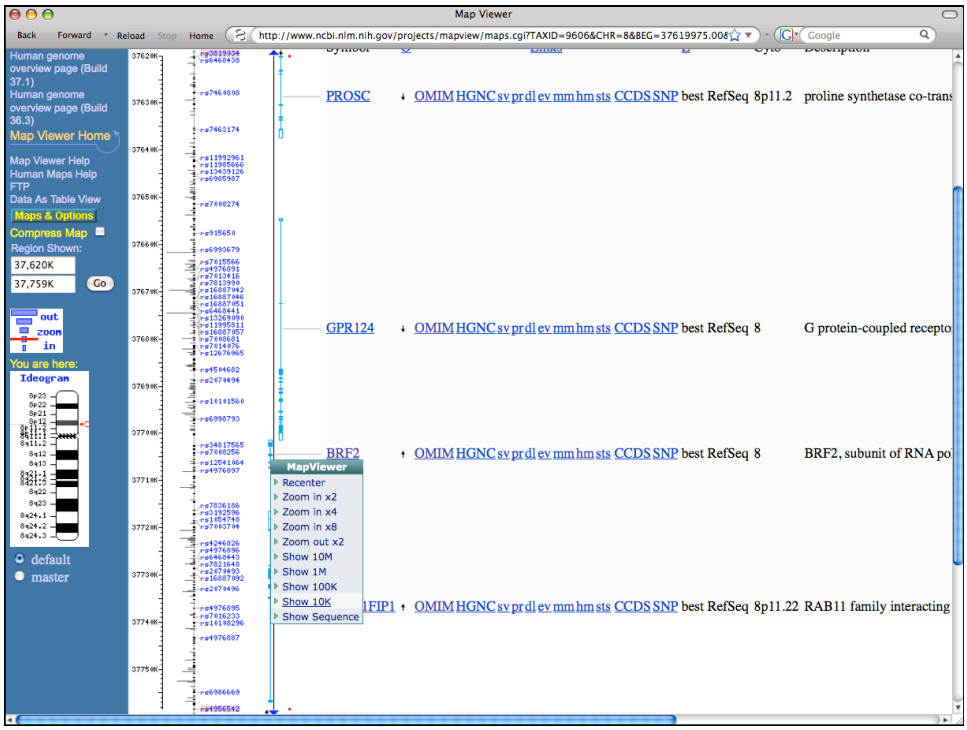

NCBI: Entrez Gene
BRF2 BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like [Homo sapiens] -
NCBI Reference Sequences (RefSeq)
RefSeqs maintained independently of Annotated Genomes
These reference sequences exist independently of genome builds. Explain
mRNA and Protein(s)
1. NM_018310.2 → NP_060780.2 RNA polymerase III transcription initiation factor BRF2
Source sequence(s) AF206673,AK001914
Consensus CDS CCDS6098.1
UniProtKB/Swiss-Prot Q9HAW0
Related Ensembl ENSP00000220659, ENST00000220659
Conserved Domains (2) summary
COG1405 Location:5 – 233 Blast Score: 252 SUA7; Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB [Transcription]
pfam08271 Location:6 – 37 Blast Score: 112 TFIIB_Zn_Ribbon; TFIIB zinc-binding
Related Sequences
Nucleotide
Protein
genomic AC130304.12 (90409..96223) None
genomic AC138356.3 None
genomic CH471080.2 EAW63351.1
EAW63352.1
EAW63353.1
mRNA AF130058.1 AAG35486.1
mRNA AF206673.2 AAG35669.2
mRNA AF298153.1 AAG30222.1
mRNA AK001914.1 BAA91975.1
mRNA AK294337.1 BAG57607.1
mRNA AK315420.1 BAG37809.1
mRNA BC010648.1 AAH10648.1
mRNA CR591369.1 None
mRNA CR602291.1 None
mRNA CR604537.1 None
mRNA CR607467.1 None
mRNA CR613791.1 None

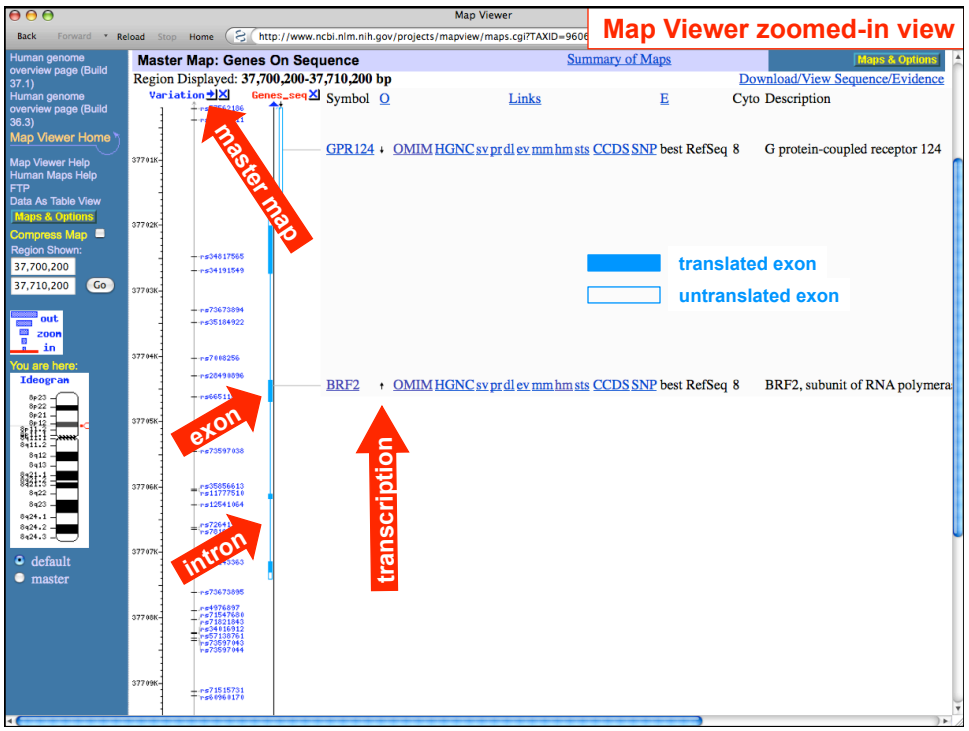

NCBI: OMIM
OMIM – BRF2 SUBUNIT OF RNA POLYMERASE III TRANSCRIPTION INITIATION FACTOR; BRF2
http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=607013
NCBI
OMIM
Online Mendelian Inheritance in Man
Johns Hopkins University
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure PMC OMIM
Search OMIM for Go Clear
Limits Preview/Index History Clipboard Details
Display Detailed Show 20 Send to
MIM *607013
Cloning
Gene Function
Mapping
References
Contributors
Creation Date
Edit History
Gene map
Entrez Gene
Nomenclature
RefSeq
GenBank
Protein
UniGene
LinkOut
komp
MGI
*607013
Links
BRF2 SUBUNIT OF RNA POLYMERASE III TRANSCRIPTION INITIATION FACTOR; BRF2
Alternative titles; symbols
BRFU
TRANSCRIPTION FACTOR IIB-RELATED FACTOR; TFIIIB50
Gene map locus 8p11.23
TEXT
CLONING
By database searching with the sequence of BRF1 (604902) as query, Schramm et al. (2000) identified BRF2, which they called BRFU, and cloned the corresponding cDNA from a total HeLa cell cDNA library. BRFU encodes a 419-amino acid protein containing a zinc ribbon domain and a core domain sharing significant homology with TFIIB (189963) and BRF1.
By peptide sequencing, Teichmann et al. (2000) identified BRF2, which they designated TFIIIB50, as a 50-kD protein that could be eluted with several other proteins from anti-TFIIIB90 (BRF1; 604902) immunoprecipitates. Sequence analysis revealed that BRF2 shares 20% identity with TFIIB and BRF1 within its N-terminal 231 amino acids.
GENE FUNCTION
Schramm et al. (2000) found that antibodies directed against BRFU were inhibitory in an in vitro translation assay dependent upon the U6 promoter. Cabart and Murphy (2001) provided evidence that BRFU is involved in the nucleation of a polymerase III-specific small nuclear RNA (snRNA) transcription initiation complex. Using recombinant proteins and HeLa cell nuclear extracts in GST pull-down and electrophoretic gel mobility assays, they determined that BRFU does not directly bind DNA; rather, it interacts with the TATA-binding protein (TBP; 600075), and these then bind cooperatively to the polymerase III promoter region of snRNA. The TBP/BRFU complex does not appear to have strict requirements for sequence motifs outside the TATA box. With use of deletion and point mutations, they determined that arg235 and phe250 of TBP, and repeat 2 of the BRFU core domain, are necessary for complex formation with DNA. BRFU also forms a stable complex on the TATA box of the polymerase II-specific adenovirus major late promoter.

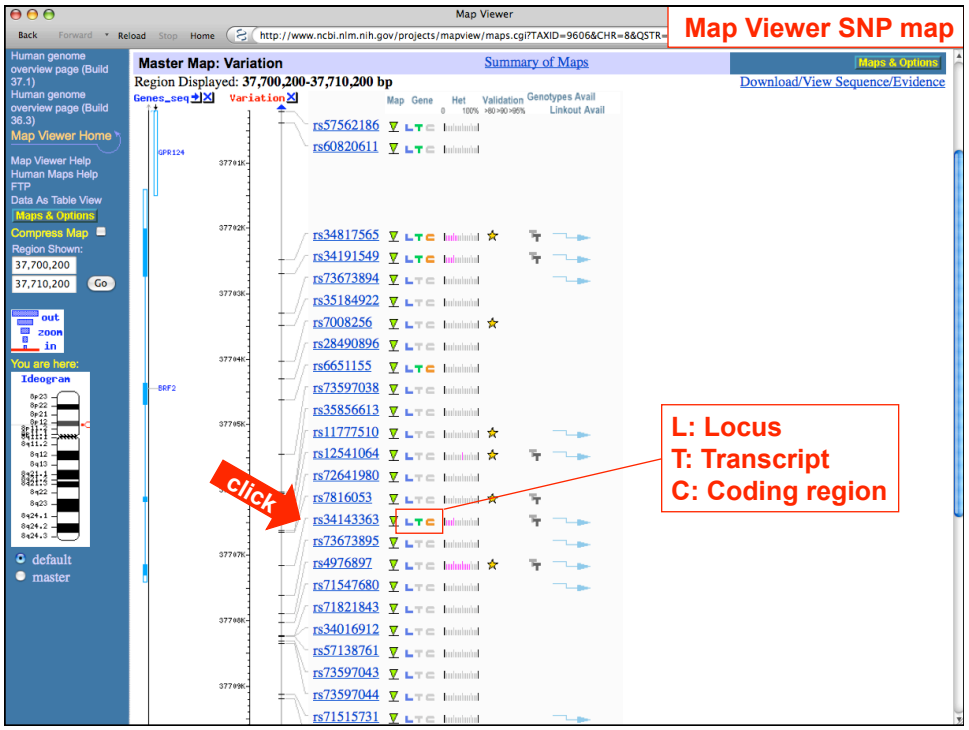

NCBI: HomoloGene (hm)
HomoloGene Result
http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Link&LinkName=gen
Search HomoloGene for Go Clear
Limits Preview/Index History Clipboard Details
Display HomoloGene Show 20 Send to
All: 1 Fungi: 0 Mammals: 0
1: HomoloGene:10127. Gene conserved in Euteleostomi
Download , Links
Genes
Genes identified as putative homologs of one another during the construction of HomoloGene.
BRF2, Homo sapiens
BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
BRF2, Pan troglodytes
BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
BRF2, Canis lupus familiaris
BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
BRF2, Bos taurus
BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
Brf2, Mus musculus
BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
Brf2, Rattus norvegicus
BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
BRF2, Gallus gallus
BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
zgc:100856, Danio rerio
zgc:100856
Proteins
Proteins used in sequence comparisons and their conserved domain architectures.
NP_060780.2
419 aa
XP_001169914.1
419 aa
XP_850010.1
421 aa
NP_001015582.1
421 aa
NP_079962.1
420 aa
NP_001019944.1
416 aa
XP_424383.1
415 aa
NP_001003536.1
423 aa
Protein Alignments
Protein multiple alignment, pairwise similarity scores and evolutionary distances.
Show Multiple Alignment
Show Pairwise Alignment Scores
Conserved Domains
Conserved Domains from CDD found in protein sequences by rpsblast searching.
TFIIB_Zn_Ribbon (pfam08271)
TFIIB zinc-binding.

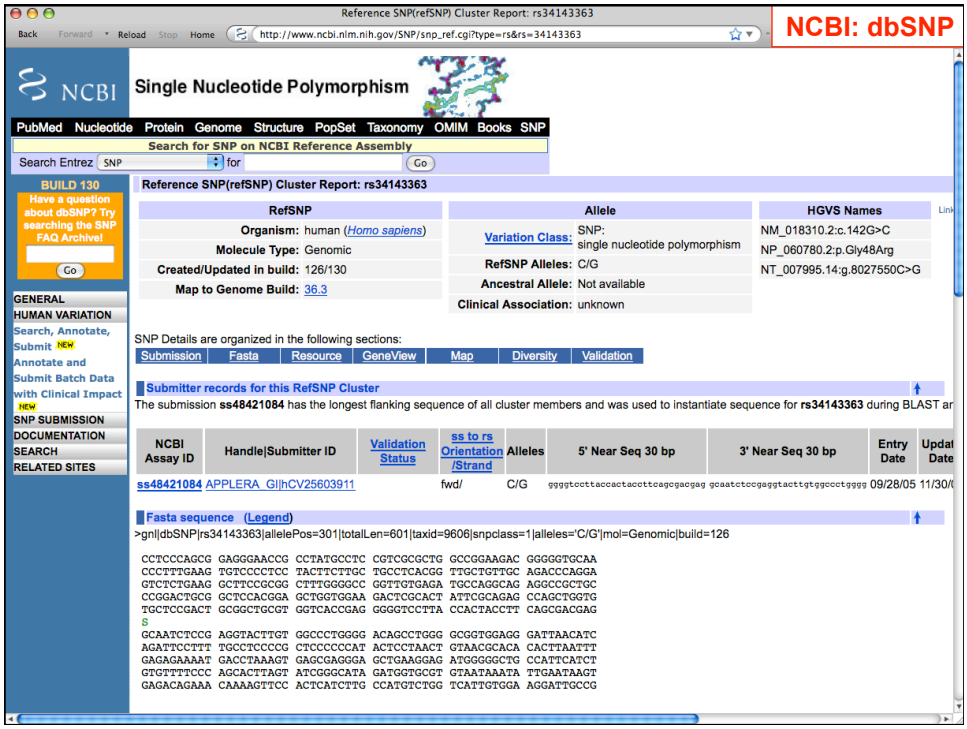
NCBI

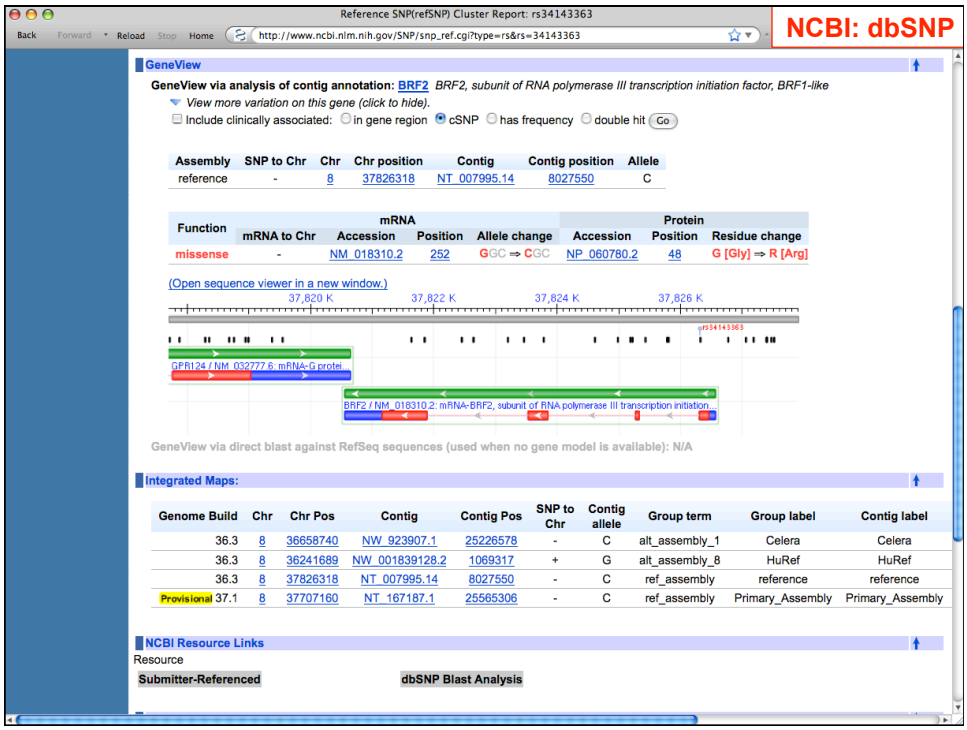
Zoom in to view greater detail




Map Viewer zoomed-in view
Master Map: Genes On Sequence
Region Displayed: 37,700,200-37,710,200 bp
master map
translated exon
untranslated exon
exon
intron
transcription
GPR124
BRF2


Map Viewer SNP map
Master Map: Variation
Region Displayed: 37,700,200-37,710,200 bp
click
L: Locus
T: Transcript
C: Coding region

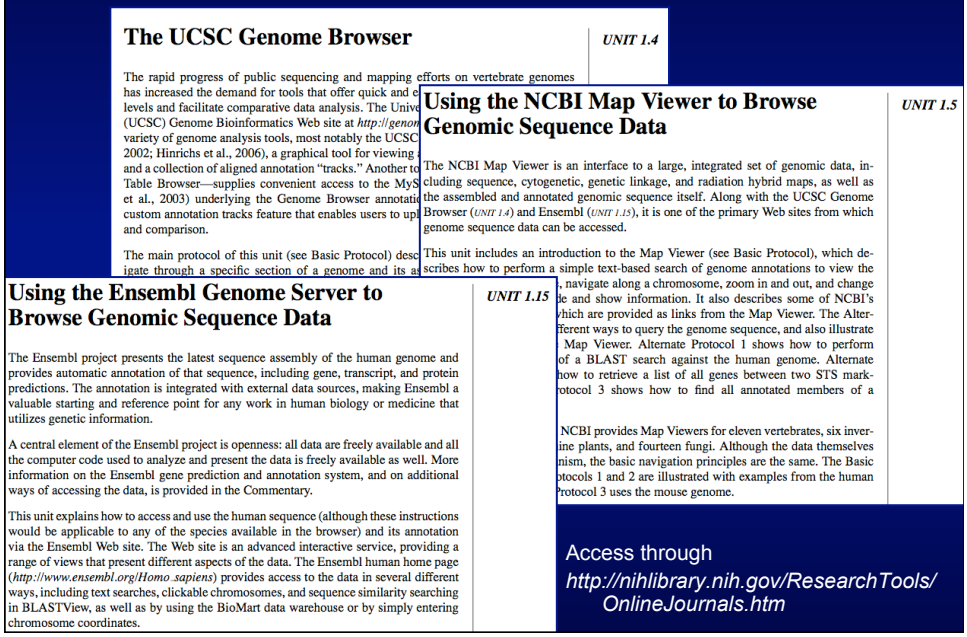

NCBI: dbSNP
Reference SNP(refSNP) Cluster Report: rs34143363
Single Nucleotide Polymorphism
Reference SNP(refSNP) Cluster Report: rs34143363
RefSNP
Allele
HGVS Names
Submitter records for this RefSNP Cluster
Fasta sequence (Legend)


NCBI: dbSNP
GeneView
Integrated Maps:
NCBI Resource Links

Additional resources

- UCSC Human Genome Browser User Guide
  http://genome.ucsc.edu/goldenPath/help/
- Ensembl Tutorials and Worked Examples
  http://www.ensembl.org/info/website/tutorials/
- NCBI MapViewer Help
  http://www.ncbi.nlm.nih.gov/mapview/static/MapViewerHelp.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 Univ (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 up 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- ferent 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- ine 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