


NATIONAL HUMAN GENOME RESEARCH INSTITUTE *Division of Intramural Research*





Current Topics in Genome Analysis 2014

Week 4: Genome-Scale Sequence Analysis

Tyra Wolfsberg, Ph.D.

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




Current Topics in Genome Analysis 2014

Tyra Wolfsberg, Ph.D.

*No Relevant Financial Relationships with
Commercial Interests*

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research



Graphical Genome Browsers

- UCSC Genome Browser
<http://genome.ucsc.edu>
- Ensembl
<http://www.ensembl.org>

Web-based access to genome data

- BioMart
<http://www.ensembl.org/biomart>
- Galaxy
<https://usegalaxy.org>



Types of data integrated in genome browsers

- Same starting material for all genome browsers: genomic sequence
- Annotations calculated independently by each genome browser
 - Genes
 - RefSeq mRNAs (non-redundant)
 - GenBank mRNAs (redundant)
 - ESTs
 - Gene predictions
 - SNPs
 - Non-coding functional elements



Genome Sequence Assemblies

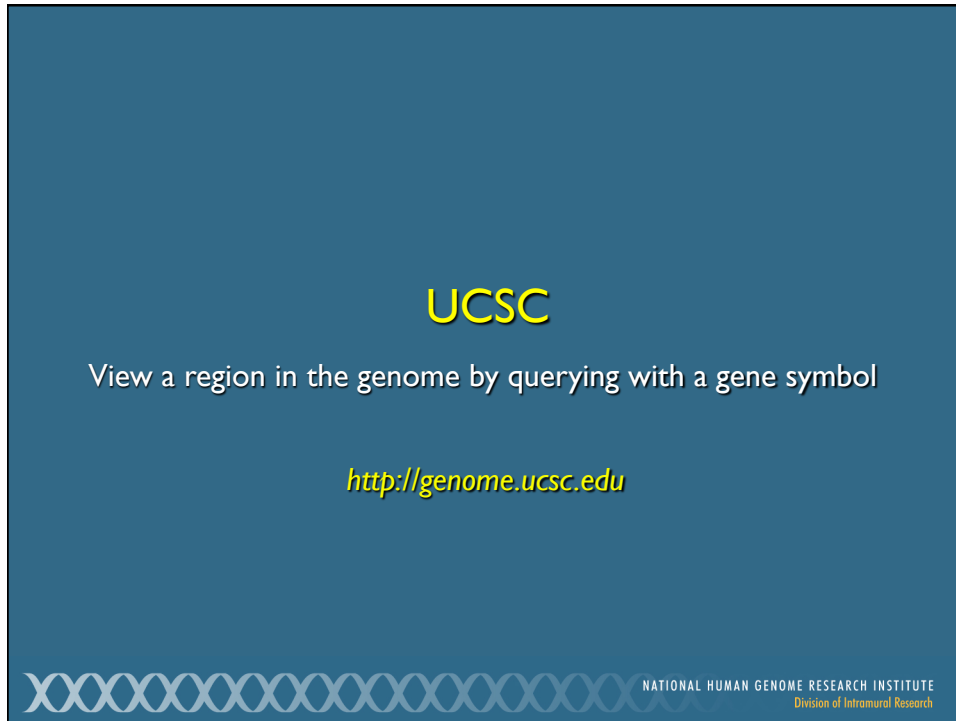
- Complex algorithms needed to incorporate all sequence data
- Assemblies updated periodically as new sequence becomes available
 - Mouse, human, and zebrafish (future) genomes assembled by the Genome Reference Consortium (GRC)
 - Other genomes assembled by sequencing centers or consortia
- Updated assemblies not available immediately in the Genome Browsers
 - “Pre-release” assemblies and annotations
 - UCSC: <http://genome-preview.cse.ucsc.edu/>
 - pre!Ensembl: <http://pre.ensembl.org/>
 - UCSC and Ensembl provide archive of old assemblies
- IF YOU ARE COMPARING DATA FROM DIFFERENT GENOME BROWSERS, MAKE SURE YOU ARE LOOKING AT THE SAME VERSION OF THE ASSEMBLY



GRCh38 human genome assembly

- First new human genome assembly in 4 years
 - Released in December, 2013
 - Fixes sequences and misassembled regions
 - Fills or reduces gaps
- 261 alternate loci
 - Many from LRC/KIR area of chr19 and MHC region on chr6
 - Example: chr6_GL000250v2_alt
- Preliminary browser (hg38) available at UCSC
- Posted on pre!Ensembl later in March
- UCSC’s liftOver converts coordinates between assemblies



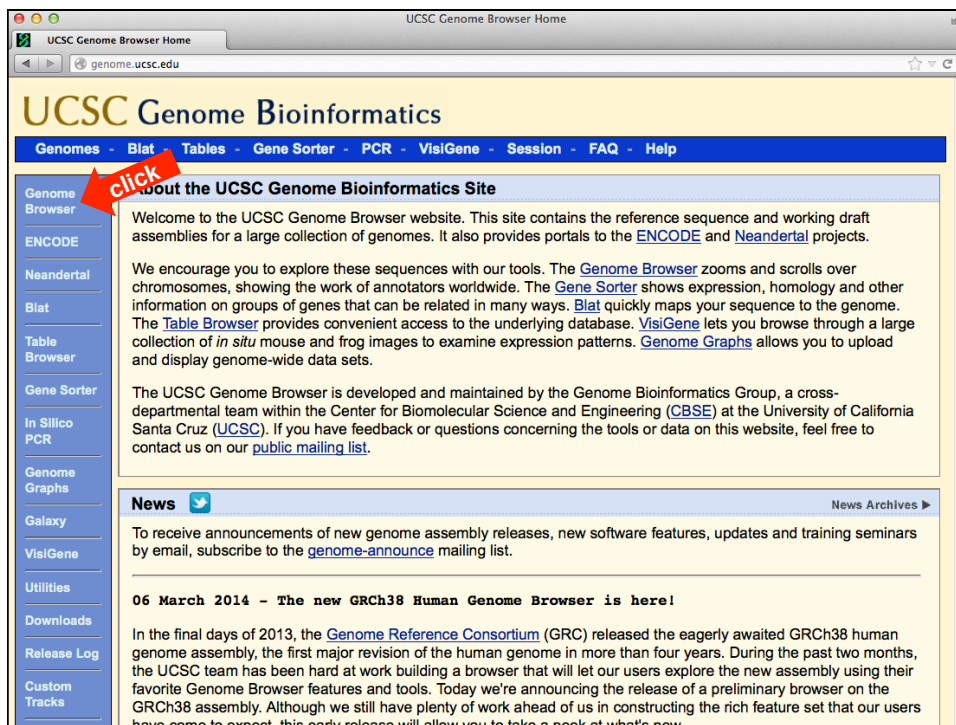


UCSC

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

<http://genome.ucsc.edu>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research



UCSC Genome Browser Home

genome.ucsc.edu

UCSC Genome Bioinformatics

Genomes - Blat - Tables - Gene Sorter - PCR - VisiGene - Session - FAQ - Help

click

About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to the [ENCODE](#) and [Neanderthal](#) projects.

We encourage you to explore these sequences with our tools. The [Genome Browser](#) zooms and scrolls over chromosomes, showing the work of annotators worldwide. The [Gene Sorter](#) shows expression, homology and other information on groups of genes that can be related in many ways. [Blat](#) quickly maps your sequence to the genome. The [Table Browser](#) provides convenient access to the underlying database. [VisiGene](#) lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. [Genome Graphs](#) allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE) at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our [public mailing list](#).

News

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [genome-announce](#) mailing list.

06 March 2014 - The new GRCh38 Human Genome Browser is here!

In the final days of 2013, the [Genome Reference Consortium](#) (GRC) released the eagerly awaited GRCh38 human genome assembly, the first major revision of the human genome in more than four years. During the past two months, the UCSC team has been hard at work building a browser that will let our users explore the new assembly using their favorite Genome Browser features and tools. Today we're announcing the release of a preliminary browser on the GRCh38 assembly. Although we still have plenty of work ahead of us in constructing the rich feature set that our users have come to expect, this early release will allow you to take a peek at what's new.

The screenshot shows the UCSC Genome Browser Gateway search page. The search criteria are: group: Mammal, genome: Human, assembly: Feb. 2009 (GRCh37/hg19), position: chr8:39601255-39695808, and search term: adam2. A dropdown menu shows search results for ADAM2 and its various transcript variants and pseudogenes. Below the search results, there is a section titled "Human Genome Browser - hg19 assembly (sequences)" with a paragraph explaining the February 2009 human reference sequence (GRCh37) and a "Sample position queries" section. A table lists "Request" and "Genome Browser Response" for various queries like chr7, chrUn_gl000212, 20p13, chr3:1-1000000, and chr3:1000000+2000.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
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

The screenshot displays a detailed genomic track for the region chr8:39,601,255-39,695,808. The top track shows the gene structure for ADAM2, with exons and introns clearly labeled. Below this are various tracks including RefSeq Genes, Human mRNAs, Spliced ESTs, Layered H3K27Ac, DNase Clusters, Tbx1 Factor ChIP, 100 Vert. Cons., Rhesus Mouse Dog Elephant, and Common SNPs. A red arrow points to the ADAM2 gene track, and another red arrow points to an intron. The interface includes navigation controls like "zoom in" and "zoom out", and a "New! On-site workshops" link.

Human Gene ADAM2 (uc003xnj.4) Description and Page Index

UCSC Gene details

Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

Human Gene ADAM2 (uc003xnj.4) Description and Page Index

Description: Homo sapiens ADAM metallopeptidase domain 2 (ADAM2), transcript variant 1, mRNA.
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. The encoded protein is a subunit of an integral sperm membrane glycoprotein called fertilin, which plays an important role in sperm-egg interactions. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, May 2013].

Transcript (Including UTRs)
 Position: chr8:39,601,255-39,695,808 Size: 94,554 Total Exon Count: 21 Strand: -

Coding Region
 Position: chr8:39,602,379-39,695,704 Size: 93,326 Coding Exon Count: 20

Page Index	Sequence and Links	UniProtKB Comments	CTD	Gene Alleles	Microarray
RNA Structure	Protein Structure	Other Species	GO Annotations	mRNA Descriptions	Other Names
Model Information	Methods				

Data last updated: 2013-06-14

GNF Expression Atlas 1 Human Data on Affy U95 Chips

Human chr8:39,601,255-39,695,808 - UCSC Genom

UCSC Change Track Display

UCSC Genes (RefSeq, GenBank, CCDS, RefSeq, tRNAs & Comparative Genomics)

RefSeq Genes

Sequences

Human mRNAs

Spliced ESTs

Layered H3K27ac

DNAse Clusters

Txn Factor ChIP

100 Vert. Cons

100 Vert. Cons

Multiple Alignments of 100 Vertebrates

Common SNPs (136)

RepeatMasker

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse 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. expand all

Mapping and Sequencing refresh

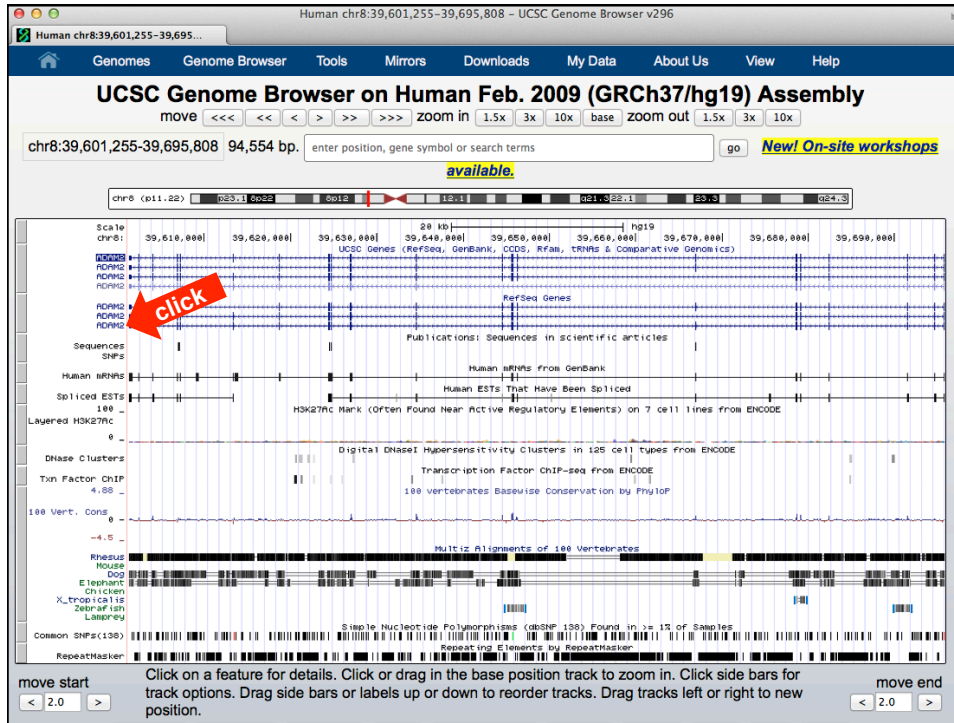
Genes and Gene Predictions refresh

UCSC Genes RefSeq Genes AceView Genes CCDS Ensembl Genes EvoFold

pack dense hide Geneid Genes Genscan Genes H-Inv 7.0 IKMC Genes Mapped

Exoniphy dense squish pack

click



RefSeq Gene ADAM2 UCSC RefSeq Gene details

RefSeq: [NM_001464.4](#) **Status:** Reviewed
Description: Homo sapiens ADAM metallopeptidase domain 2 (ADAM2), transcript variant 1, mRNA.
CCDS: [CCDS34884.1](#)
CDS: 3' complete
OMIM: [601533](#)
Entrez Gene: [2515](#)
PubMed on Gene: [ADAM2](#)
PubMed on Product: [disintegrin and metalloproteinase domain-containing protein 2 isoform 1 preproprotein](#)
GeneCards: [ADAM2](#)
AceView: [ADAM2](#)

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. The encoded protein is a subunit of an integral sperm membrane glycoprotein called fertilin, which plays an important role in sperm-egg interactions. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, May 2013].

mRNA/Genomic Alignments

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL	
browser		2671	100.0%	8	-	39601255	39695808	NM_001464	1	2671	2686

- [mRNA Sequence](#) (mRNA derived from the genomic sequence)
- [Genomic Sequence](#) from the assembly
- [CDS FASTA alignment](#) from multiple alignment

click

UCSC RefSeq Gene details

Genomic Sequence Near Gene

Get Genomic Sequence Near Gene

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

Sequence Retrieval Region Options:

- Promoter/Upstream by bases
- 5' UTR Exons
- CDS Exons
- 3' UTR Exons
- Introns
- Downstream by bases
- One FASTA record per gene.
- One FASTA record per region (exon, intron, etc.) with bases
- 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, avoid extending past the edge of the chromosome.

Sequence Formatting Options:

- Exons in upper case, everything else in lower case.
- CDS in upper case, UTR in lower case.
- All upper case.
- All lower case.
- Mask repeats: to lower case to N

1000 nt upstream of ADAM2

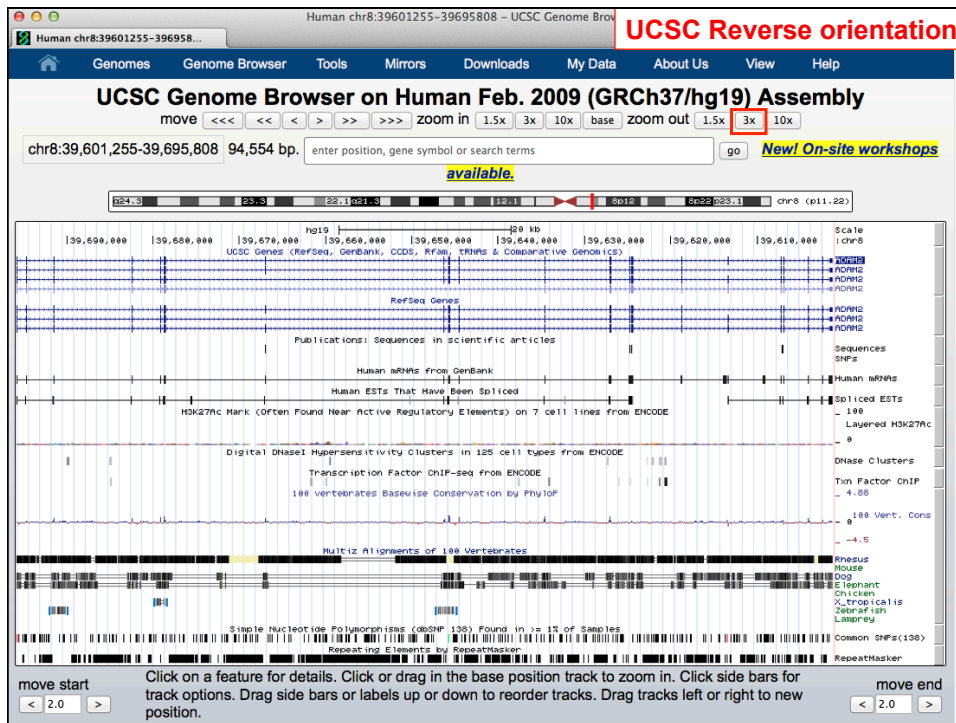
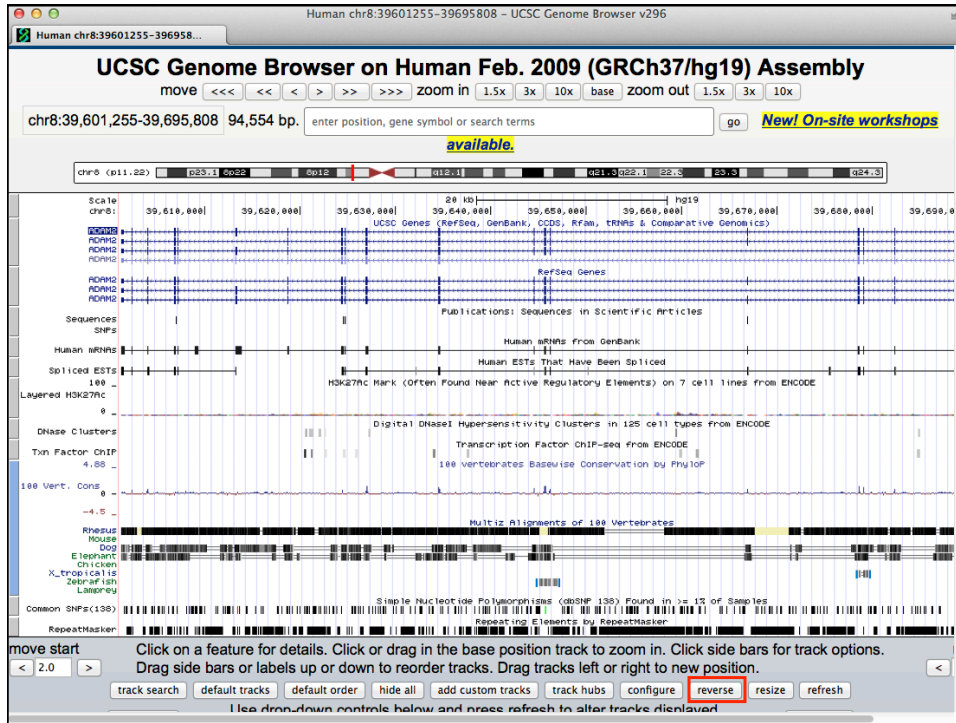
```
>hg19_refGene_NM_001464 range=chr8:39695809-39696808 5'pad=0 3'pad=0
taaattagtagctcttggaaaactgtcggaagatcttaccacacata
ccctgtgatcgcgaacttcaactctagaatacaoaagtagaactcctt
acttattacaccaaaaagcatgagaagaatgtttatagctaaatattat
tttaaatagctggaacaataaacaacaaatattcaataacagtaaaat
ggaacacaagtggtgttatatttaagaattgtaataacaccaatga
ggataaaacagaactattgcttggatgaaccttacaatcactctattaa
aagaaccagacatgaaagatgagatggtgatgctctacttgcgaaaa
gtcaaaaacagacaaaagaatcttggttgttagaagctatggtgtg
gaggttggaactcggggatttgggtgttcttttcaattctctcatctg
gtactagttaacgtgttttttttccacttgaatttaattgacctgtg
aacttatgatttataataacttttttctgtttttgtttgttctttttt
ttttttttttttttttttttgacggaattctgctctctcaaccaggtg
gagtgcaaggtaaggtcttttctgcaactgcaacctctgctctaggt
taaacgcatctctgctcagcttcccgagtagctgggattcaggcac
ccgccacatgctcgtcaattttttttgtatttttagtaacagagggg
tttcaccatgttgcggaactggtctcgaactcctgatacgtgtttatat
atatcaattgaaatttacttaagaaggtttataaattctctgttcc
taagctgttggaaagtattttgtgtgtgtgtgacttaattaggatca
cgtccagtgagtgctgtctcgaagagacagggctcaggagctcagcac
gttccacagcaccacacccaacctcagcccacctgggctctccagcc
```

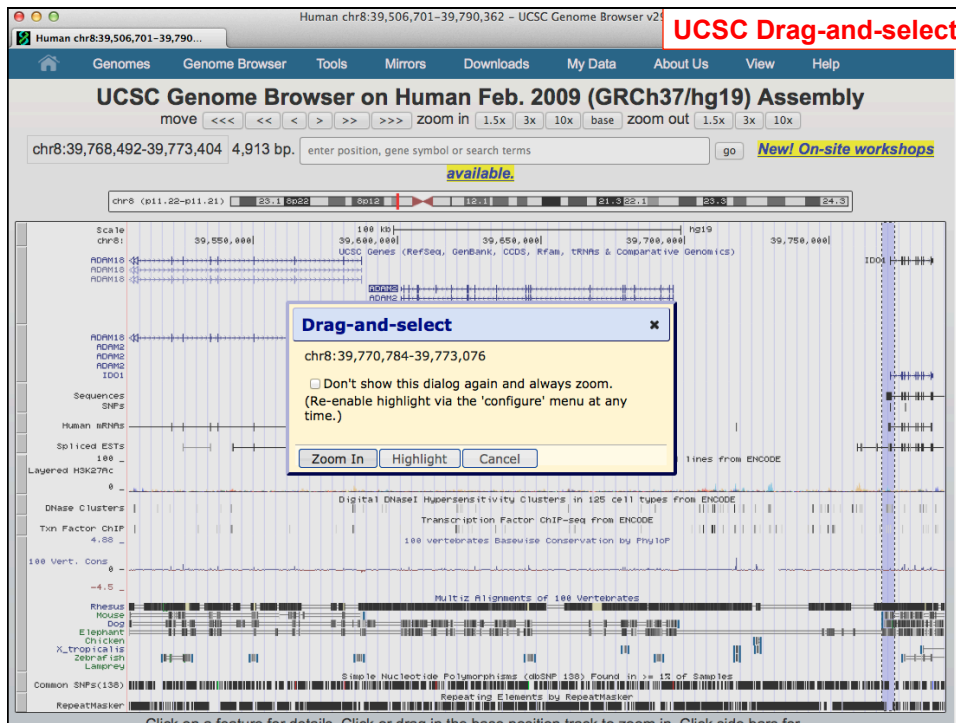
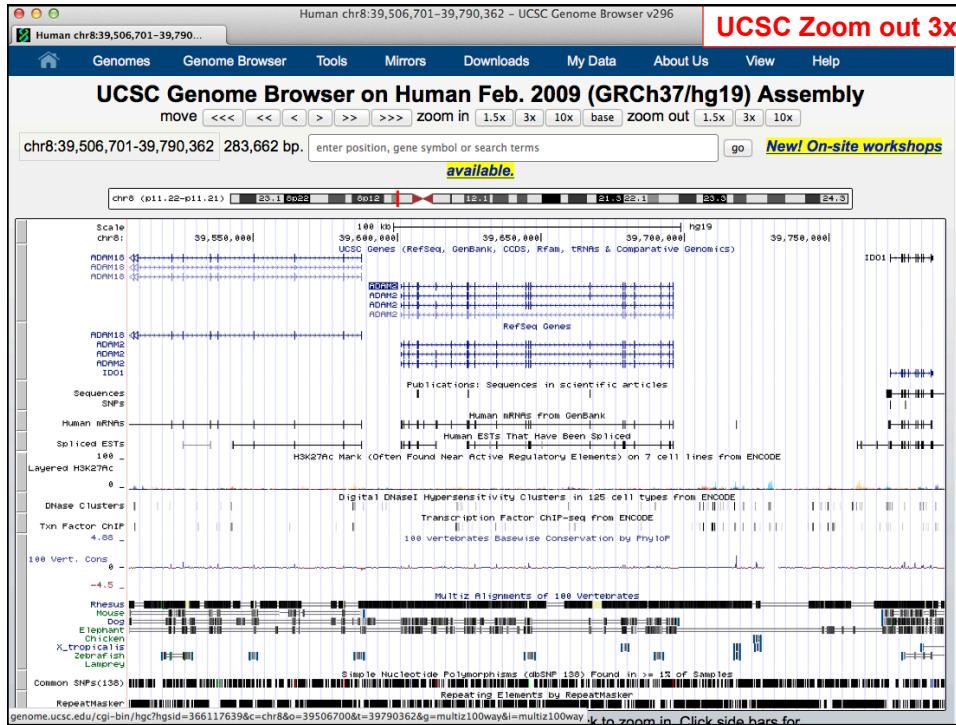
UCSC

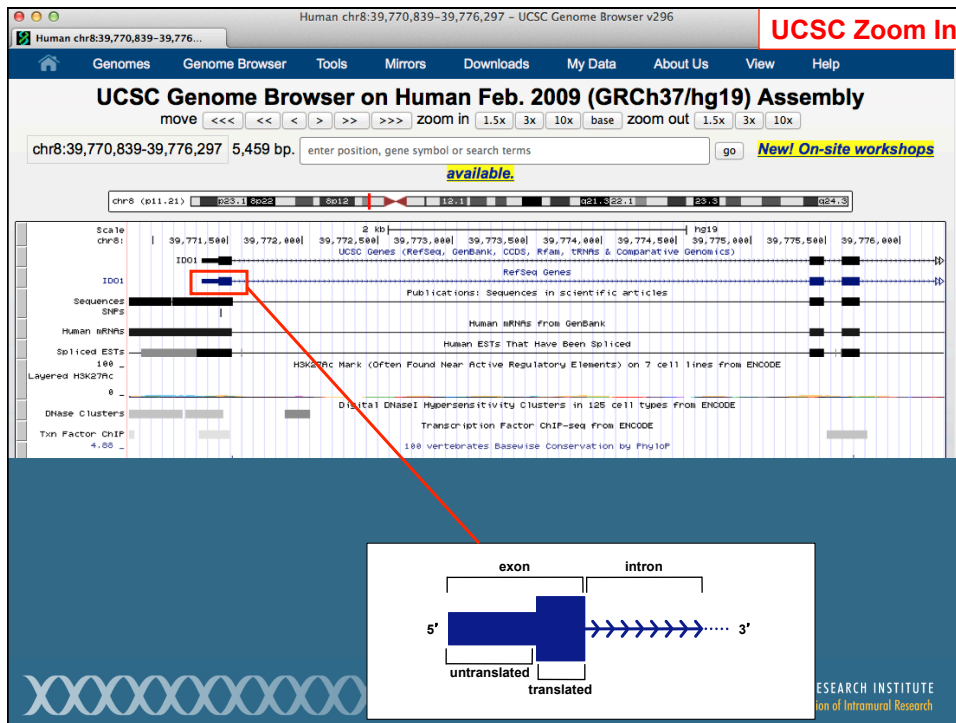
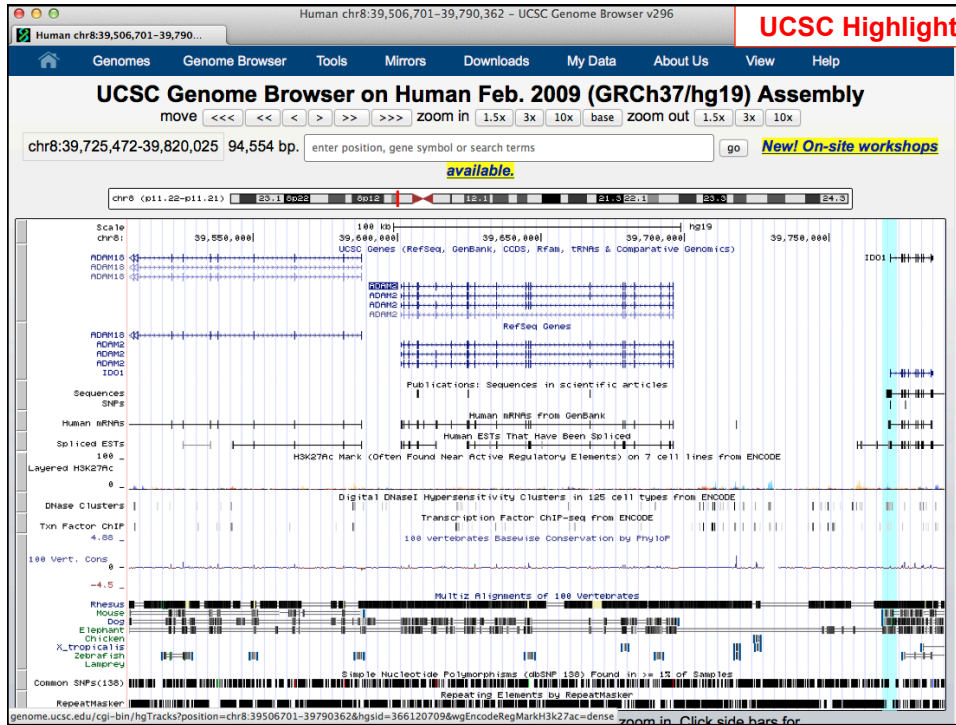
Navigating around the Genome Browser

<http://genome.ucsc.edu>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research








UCSC

Configure Track on the Genome Browser

<http://genome.ucsc.edu>



NATIONAL HUMAN GENOME RESEARCH INSTITUTE
 Division of Intramural Research

Human chr8:39601255-39695808 - UCSC Genome Browser v296

Human chr8:39601255-39695808

hide hide hide

Stanf Nucleosome SUNY SwitchGear SwitchGear TSS TFBS Conserved TS miRNA sites UCSF Brain Methyl

hide hide hide hide hide

UMMS Brain Hist UW Repli-seq Vista Enhancers

hide hide hide

Comparative Genomics refresh

Conservation full MmCf Cons 46-Way hide Cons Indels hide Evo CpG hide GERP hide phastBias gBGC hide

Primate Chain/Net Placental Chain/Net Vertebrate Chain/Net

hide hide hide

Neandertal Assembly and Analysis refresh

Denisova Assembly and Analysis refresh

Variation refresh

Common SNPs(138) dense 1000G Ph1 Accsbl hide 1000G Ph1 Vars hide All SNPs(135) hide All SNPs(137) hide All SNPs(138) hide

Common SNPs(135) hide Common SNPs(137) hide DGV Struct Var hide Flagged SNPs(135) hide Flagged SNPs(137) hide Flagged SNPs(138) hide

Genome Variants hide GIS DNA PET hide HAIB Genotype hide HapMap SNPs hide HGDP Allele Freq hide Mult. SNPs(135) hide

Mult. SNPs(137) hide Mult. SNPs(138) hide NumtS Sequence hide Segmental Dups hide Self Chain hide SNP/CNV Arrays hide

Repeats refresh

RepeatMasker Interrupted Rpts Microsatellite Simple Repeats

dense hide hide hide

refresh

Common SNPs(138) Track Settings

UCSC SNP Track details

Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

Common SNPs(138) Track Settings

Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples (All Variation tracks)

Display mode: **pack** Submit

Include Chimp state and observed human alleles in name:
 (If enabled, chimp allele is displayed first, then '>', then human alleles).

Use Gene Tracks for Functional Annotation

Filtering Options

Coloring Options

SNP Feature for Color Specification: Function Set defaults

The selected "Feature for Color Specification" above has the selection of colors below for each attribute. Only the color options for the feature selected above will be used to color items; color options for other features will not be shown. If a SNP has more than one of these attributes, the stronger color will override the weaker color. The order of colors, from strongest to weakest, is red, green, blue, gray, and black.

Unknown **black** Locus **black** Coding - Synonymous **green** Coding - Non-Synonymous **red**
 Untranslated **black** Intron **black** Splice Site **black**

[View table schema](#)
 Data last updated: 2013-10-22

Description

This track contains information about a subset of the single nucleotide polymorphisms and small insertions and deletions (indels) —

Human chr8:39601255-39695808 - UCSC Genome Browser v296

UCSC SNP Track

ADPRK2 RefSeq Genes
 Hedenstierna2009 Sequences in Articles: PubMedCentral and Elsevier
 Fuluck1997 Chen1999 SNPs in Publications
 Human mRNAs Human mRNAs from GenBank
 Human ESTs That Have Been Spliced
 Spliced ESTs HSK27c Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
 Layered HSK27c
 DNase Clusters Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE
 Tbx Factor ChIP Transcription Factor ChIP-seq from ENCODE
 100 Vert. cons. 100 vertebrates: Basewise Conservation by PhyloP
 Rhesus Mouse
 Elephant
 Chicken
 X_trogl (ca) 12
 Zebra finch
 Lampbrush

Green: coding-synonymous
Red: coding-nonsynonymous
Black: other SNPs

non-synonymous SNP

synonymous SNP

Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples

r334966682	r561832	r433959	r282012	r282095	r12681344	r79122716	r14992990	r26474177	r21449799	r72642845	r72673868
r3122080	r4073382	r433959	r282012	r11890156	r72765999	r2122091	r2974271	r7468279	r70148791	r20786938	r2113871
r11202783	r420351	r7817718	r3353758	r282083	r1947319	r1451741	r896472	r18498917	r35963524	r1176444	r73673829
r143373254	r417869	r7814938	r10996756	r2293081	r76659751	r1981386	r28			r7463449	r249796528
r119934991	r372999	r453419	r76776344	r1981387	r13265483	r2122099	r73			r7463516	r68511219
r1182534	r33664556	r1375160	r282018	r74653284	r1343664	r8591036	r8			r37420355	r2984571
r281977	r9995209	r12546338	r2122092	r13255131	r3779718	r5993989	r7			r56251288	r7844819
r281976	r79433946	r11808920	r282099	r25091141	r6987114	r14549488	r1			r446668	r34900519
r14796549	r12541881	r59233181	r7898929	r57359665	r7822312	r7387353					
r142876623	r710837	r118118518	r282088	r117681926	r7486936	r7306549					
r281975	r77164895	r118185783	r282097	r11869625	r14889	r15124433					
r7813282	r4733922	r11678387	r282086	r74077196	r1451745	r1165879					
r281974	r336539	r2544857	r3591469	r11894349	r113541240	r1181114					
r281973	r33935433	r282015	r37567475	r13253073	r59621140	r35913294	r17359980	r116226414	r117115990	r13285763	
r281972	r2388662	r220014	r282094	r14552944	r1451744	r88431931	r79582812	r116831873	r111449616		
r281971	r2318661	r7836159	r7468772	r24336527	r116778814	r114543667	r14533284	r115784084	r113345234		
r142717481	r12545821	r11828013	r11898578	r118186478	r288932198	r73615918	r141420786	r141420786	r10188537	r147837515	
r11898783	r28202873	r7579883	r28188289	r7372075	r133788299	r28118634	r5987296	r1899148	r117689148		
r118933	r18993751	r11777615	r11895856	r351297878	r11811896	r149927514	r737432885	r62511289	r115653795		
r281970	r11897516	r111044793	r2358463	r11282246	r112272578	r141182407	r141520783	r141520783	r151663795		
	r11877294	r111257911	r78222544	r288864576	r117896993	r281288834	r12677911	r117484596			
r77893623	r281969	r114747845	r112473129	r113337489	r76718899	r148683951	r14839818	r11929945	r114647659		
	r281968	r11546836	r72042841	r11781121	r1143561398	r7488889	r116226414	r114532386	r114532386		
	r59		r64467	r14785895	r88836244		r144187724	r149322982	r145734821		
	r398		294958	r72042843	r12785197		r11585778	r1149823851	r281865185		
r1478			7836214	r76387989	r146791755		r11815392		r13253377		
r1468			9959638	r28564328	r78119292		r12674643		r79478866		
r146			r14648394	r13951352	r73915965		r11764782		r11523225		
r283				r18857486	r75557624		r112717748		r61429382		
r142				r304417912	r111718285		r113274185		r11739521		
					r75598227		r151288279		r113539436		

UCSC
ENCODE tracks

<http://genome.ucsc.edu>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Human chr8:39663367-39784211 - UCSC Genome Browser v296

ENCODE tracks

Sequences
hide

mRNA and EST refresh

Human mRNAs Spliced ESTs CGAP SAGE Gene Bounds H-Inv Human ESTs

Integrated Regulation from ENCODE Tracks (All Regulation tracks)

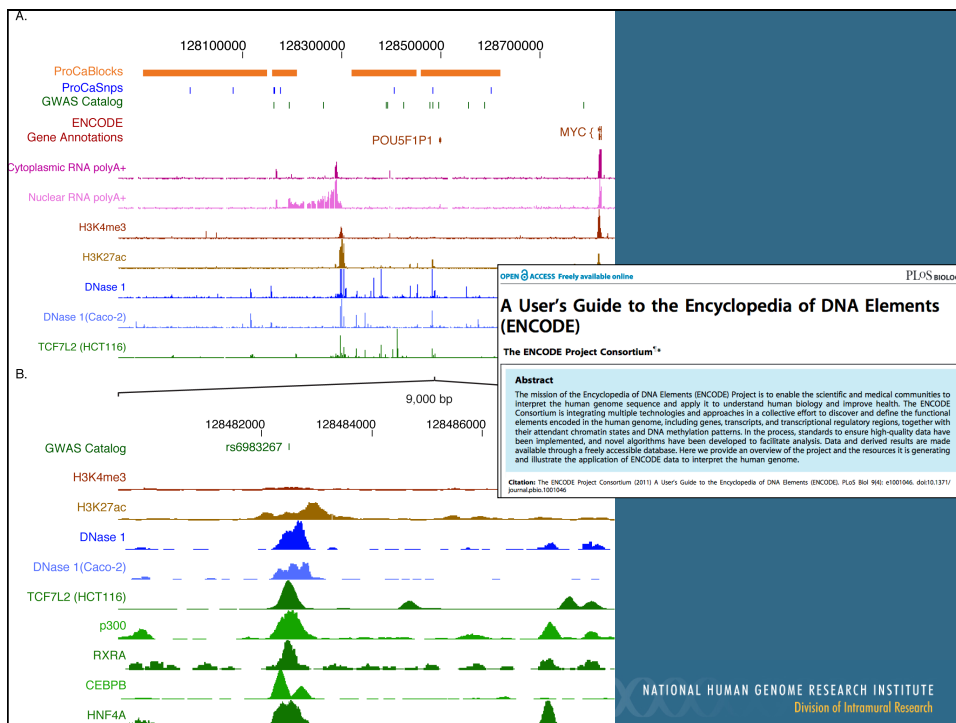
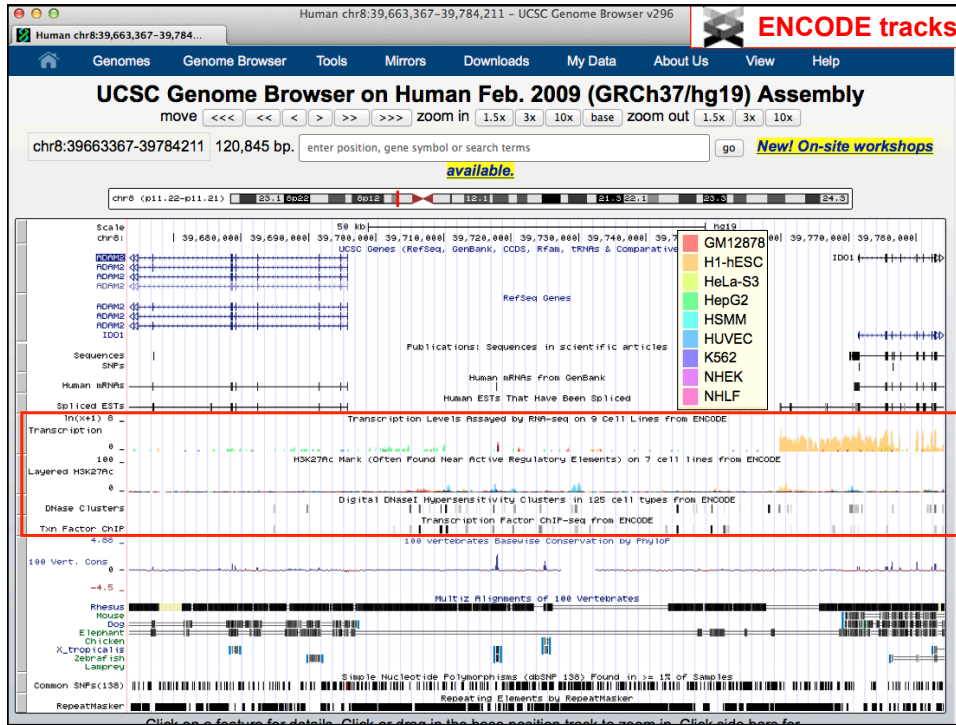
Display mode: show Submit

- All
- Transcription Transcription Levels Assayed by RNA-seq on 9 Cell Lines from ENCODE
- Layered H3K4Me1 H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
- Layered H3K4Me3 H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
- Layered H3K27Ac H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
- DNase Clusters Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE
- DNase Clusters V1 Digital DNaseI Hypersensitivity Clusters in 74 cell types (2 reps) from ENCODE
- Txn Fac ChIP V3 Transcription Factor ChIP-seq Clusters V3 (161 targets, 189 antibodies) from ENCODE ENCODE Mar 2012 Freeze
- Txn Factor ChIP Transcription Factor ChIP-seq from ENCODE

Regulation refresh

- ENCODE Regulation... show
- CD34 Dnase1
- CpG Islands
- ENC Chromatin...
- ENC DNA Methy...
- ENC DNase/FAIRE...
- ENC Histone...
- ENC RNA Binding...
- ENC TF Binding...
- FSU Repli-chip
- NKI Nuc Lamina...
- ORegAnno
- Stanf Nucleosome
- SUNY SwitchGear
- SwitchGear TSS
- TFBS Conserved
- TS miRNA sites
- UCSF Brain Methyl
- UMMS Brain Hist
- UW Repli-seq
- Vista Enhancers

click



UCSC

Find a chicken homolog of a human protein

<http://genome.ucsc.edu>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

disintegrin and metalloproteinase domain-containing protein 2 isoform - Protein

NCBI Entrez Protein

Protein Search

Advanced Help

Display Settings: FASTA Send to:

disintegrin and metalloproteinase domain-containing protein 2 isoform 1 preproprotein [Homo sapiens]

NCBI Reference Sequence: NP_001455.3

[GenPept](#) [Graphics](#)

>gi|55743080|ref|NP_001455.3| disintegrin and metalloproteinase domain-containing protein 2 isoform 1 preproprotein [Homo sapiens]
MWRVFLLSGLGRLMDSNFDLSPVQITVPEKIRSIKEGIESQASYKIVIEGKPYTVNLMQRNPLPHNF
RVYSYSGTGMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPLESSVGFHVI
YQVHKKADVSLNERDLISRDLSFKLSVEPQQDFAKIEMHIVIEKQLYNHMSDPTVVAQRVFLIG
LTMALFVFNITILLSSLELWIDENKLTATGEMELLHFTLRWKTLYLVPIDVAFLLYREKSNVQGA
TFQKMKDANYAGGVVLRHTISLESIAVLAQLLSMGIYDDINKQCQSGAVICINNPFAIHFSQYKI
FSNCSPEFAHFISKQSQCLNQPRLDPPFKQAVQCNALKEAGECCDCTQDCALIGTCCDIATCR
FKAGSNCAEGPCEENCLFMSKERMRPSPFEBEDLPEYCNGSGASCPENHYVOTGHPCLNQTIDGVCN
SGDKQCTDTGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYQCEADNLQCKLICKYVGFLLQIPRA
TIIYANISGHLIAVEFASDHADSQKMIKDGTSKSGSNKVCNRCVSSSYLGYDCTTDKCNDRGVCNNK
KHCHCSASVLPDSCVQSDLWPGSIDSNGFPVVAIPARLPERRYIENIYHSPKMRWPFLLIPFFLIIFC
VLIAMVKNVFRKKWRTEYSSDEQPESESEPKG

Change region shown

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Articles about the ADAM2 gene

Testicular and epididymal ADAMs: expression and function [Nat Rev Urol. 2012]

Evolutionary divergence and functions of the ADAM and ADAMTS: [Hum Genomics. 2009]

Mapping, sequence, and expression analysis of the human fertilin beta gi [Genomics. 1997]

See all...

Identical proteins for NP_001455.3

Sequence 28430 from patent US [AHD78786]

unnamed protein product [Homo] [CBH30599]

ADAM metalloproteinase domain [EAW63273]

See all...

Pathways for the ADAM2 gene

Interaction With The Zona Pellucida

Fertilization

Chicken BLAT Search

UCSC BLAT search

Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

Chicken BLAT Search

BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

```
>q1[55743080]ref|NP_001455.3| disintegrin and metalloproteinase domain-
containing protein 2 isoform 1 precursor [Homo sapiens]
MWRLVLLSGLRMDNFSLFVQITVPEKIRSLIKEGIESQASYKIVIEGKPYIVNLMOKNLFPHNF
RVYSYSGTGMKPKLDQDFQNFCHYQYIEGPKSVVMVSTCTGLRGVLOFENVSYGLEPSSVGFPHVI
YQVHKKADVSLYNEKDIERSRDLSPKLSQVPEPQDFAKYTEMHIVIEKOLYNHMGSDPTVVAKVFLIG
LTHAIFVSNITIISSLELWIDENKLTATGEANELHTFLMRTSYLVLRPHDVAFLLVYREKSNHYGA
TFQGMCDANYAGGVLPRTISLSESLAVILAQLLSMOTTYDDINKCCGAVCIIMPEAIEFGVKI
FNSCFEDFAHFISKQKSCQLNQPRLDFPFKQAVCGNAKLEAGEECCKSTEDCALIGECCTIACR
FKAGSNCAGPCENCLFMSKERMCRPSEECDFEYCGSSASCENHYVORGHPCGLNWCIDGVCN
SGDKQCTDPTFGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYTOCADNLQCKLICKYVGFLLQIPRA
TIIYANISGELCIAVEFASDHADSQMWDKGTSCGSNKVCRNORCVSSYLYGDCYTDKCDNRGVCNNK
RHCRCASLYLPDCSVQSDLWFGGSDISGNFFVAIPARLPERATYENIYHSKPRWFFFLIFPFLIFC
VLAIAHWKVPQRKWRTEVSSDQPSSEEPKQ
```

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: No file selected.

Only DNA sequences of 25,000 or fewer bases and protein or translated sequence of 10000 or fewer letters will be processed. Up to 25 sequences can be submitted at the same time. The total limit for multiple sequence submissions is 50,000 bases or 25,000 letters.

For locating PCR primers, use [In-Silico PCR](#) for best results instead of BLAT.

About BLAT

Chicken BLAT Results

UCSC BLAT search

Genomes Genome Browser Tools Mirrors Downloads My Data About Us 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%	22	++	2453105	2453290	186

UCSC Genome Browser on Chicken Nov. 2011 (ICGSC Gallus_gallus-4.0/galGal4) Assem

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr22:2,453,105-2,453,290 186 bp. enter position, gene symbol or search terms [New! On-site workshops](#)

[available.](#)

UCSC BLAT search

Genomes Genome Browser Tools Mirrors Downloads My Data About Us 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%	22	++	2453105	2453290	186

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

```

mrvlfillg lgglrmdanf dslpvqitvp ekirsiikeg iesqasyyiv iegkpytvnl 60
mqknflphn rvysysetgi mkpldqdfgn zchygyyieg ypkavmvat ctglrgvlqf 120
envsygiepl essvgefvi yqwkhhkadv slynekdiez rdisfklqsv epqgdakyi 180
emhvivekql ynhvgsdttv vaqkvfqlig ltnaifvsfn itiillslel widenkiatt 240
geanelhtf lrwktsylvl rphdvafllv yreksnyvga tfqgkmdan yaggvvlhpr 300
tisleslavi laqllsismg ityddinkcq csgawcilmpp ealhfsyvkf fencsfedfa 360
hfiskqkeqg lhngrldpf fkgqavcuna klsaegeode qteqcalig etocciatcr 420
fkagsnaeq pccenclfms kermcrpsfe ecdlpeycng ssascpenhy vqtghpcgln 480
qwicidgvcn sgdkqctdft gkevefsgpe cyshlnsktd vsqncglids gytgceadNL 540
qCGKLICKVv gkfillqipra TIYAnisgH LClavefasd hadsqkmwik DGTsCGenKV 600
crnqrsvss ylygydcttdk cndrgvcnnk khchosaayl ppdcsvgsdl wpgsidsdn 660
fppvalparl peryieniy hskpmrwpff lfipffilfo vliainvkvn fgrkkwrted 720
yssdeqese sepkq
    
```

Chicken.chr22 :

```

AATCTGggcT GTGGAAACT CATCTGCaca TACccaaac gqgttcctt caccaaatta 2453164
aagggtACCA TCATCTATGC Tcaagtgcaa gaaCATCTGT Gcgtgtcttt t gatgtaatg 2453224
catgcacct ccgggacaga tcctctcctg gttAGGATG GCACGaaaTG CCGTccogga 2453284
AAGGTA
    
```

Side by Side Alignment*

```

0001615 N L Q C G K L I C K Y 0001647
>>>>>> | G | | | | | | | T | >>>>>>
2453105 aactctggcgtgtggaaaactcatctgcacatac 2453137
    
```

UCSC

Add your own custom tracks

<http://genome.ucsc.edu>

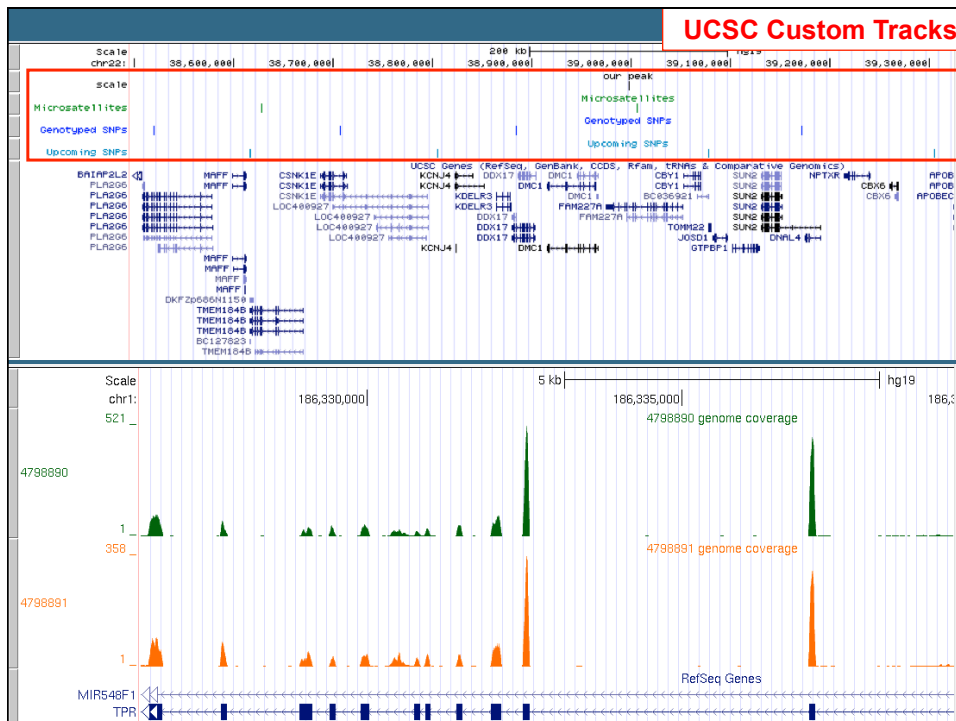
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
 Division of Intramural Research

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



Types of UCSC custom tracks

- Upload annotation data from your computer
 - Tracks viewable only from your computer
 - Discarded after 48 hours
- Post annotation data to your Web site
 - Tracks can be shared with anyone
 - Never discarded
- Create a Session with specific track combinations, including custom tracks
 - Session can be shared or non-shared
 - Session persists for 4 months; custom tracks for 48 hours

<http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html#CustomTracks>



NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

UCSC

Table Browser

<http://genome.ucsc.edu>



NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

UCSC Table Browser

- Download track in text format or create custom tracks
- Retrieve DNA sequence
 - Get sequence 200 nt upstream of each RefSeq gene
- Calculate intersections between tracks
 - List all SNPs in a RefSeq gene
- Filter track data based on certain criteria
 - Show all RefSeq genes that contain only one exon



UCSC Table Browser: 200 nt upstream of each RefSeq gene

clade: Mammal genome: Human assembly: Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Predictions track: RefSeq Genes

table: refGene

region: genome ENCODE Pilot regions position chr21:33031597-33041570

identifiers (names/accessions):

filter: create

intersection: create

correlation: create

output format: sequence Send output to Galaxy GREAT

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

file type returned: plain text gzip compressed

Sequence Retrieval Region Options:

- Promoter/Upstream by 200 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
- 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 extending past the edge of the chromosome.

Sequence Formatting Options:

- Exons in upper case
- CDS in upper case
- All upper case.
- All lower case.
- Mask repeats:

```
>hg19_refGene_NM_032291 range=chr1:66999625-66999824 5'pad=0 3'pad=0 strand++ repeatMasking=none
ggaaaggctgtgtatcttggtaacaagcgggagcgtgggtggagg
gaatgggacgggaataggttctgtgtctctccggggatctgtgca
ggagatgcaggctggctaccatgtgacgggtccaagtgaggatg
gccgagcagcggcgggtgcagctcggccagctgcccctctctccc
>hg19_refGene_NM_032785 range=chr11:50489627-50489826 5'pad=0 3'pad=0 strand-- repeatMasking=none
gtccctcaggaagcgcattcctcacaagacctggccctgtgctctag
gggtggcggagcctgctcagttgccaggatgcggtggatcggttgcggag
gagcagcgcgccggggcccaagtcocagctcagcgcggcctcgcgcgc
ggggcaggccttggagcagcctcctccagcggctccctccggcctc
>hg19_refGene_NM_001145278 range=chr1:16766967-16767166 5'pad=0 3'pad=0 strand++ repeatMasking=none
agaccacgggttacagagggtctgtccatggcggggcagggcgcttc
tttgctcggagggtgtctggagggaaggagaacctctggaggaggga
gaagcctcggagggtgcggccacgtctcttgagccgggttccagcag
agggcgcaaacagggcgtgtgagggcccgagctgcgctctagcc
```

Ensembl
Variant Effect Predictor (VEP)
<http://www.ensembl.org>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

The slide features a blue background with a white DNA double helix graphic at the bottom. The text is centered and uses a mix of white and yellow colors.

Ensembl Genome Browser

BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Search: All species for Go
e.g. BRCA2 or rat X:10000..20000 or coronary heart disease

What's New in Release 75 (February 2014)

- New VEP interface
- New 'Age of base' track for human
- New GENCODE basic renderer for human and mouse

Full details of this release
More release news on our blog →

Latest blog posts

- 11 Mar 2014: GRCh38 Assembly Mapping – Updating Coordinates in the New Human Genome
- 05 Mar 2014: A Korean ENCODE adventure
- 27 Feb 2014: Ensembl 75 has been released!

Go to Ensembl blog →

Did you know...?
It's free- take our browser workshop online!

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

Ensembl receives major funding from the Wellcome Trust. Our acknowledgements page includes a list of additional current and previous funding

The screenshot shows the Ensembl Genome Browser interface. A red arrow points to the 'Variant Effect Predictor' link in the 'What's New' section. The interface includes a search bar, navigation links, and various data visualization options.

Ensembl genome browser 75: Homo sapiens – Variant Effect Predictor – Tools

Ensembl: VEP tyra@nhgri.nih.gov

Human (GRCh37) Jobs

Variant Effect Predictor

New VEP job:

Input

Species: Human (Homo sapiens)
 Assembly: GRCh37

Name for this data (optional): ADAM2 final

Input file format (details): Variant identifiers

Either paste data:

```
rs35935433
rs146646998
rs145143599
rs34417912
```

Or upload file: Browse... No file selected.

Or provide file URL:

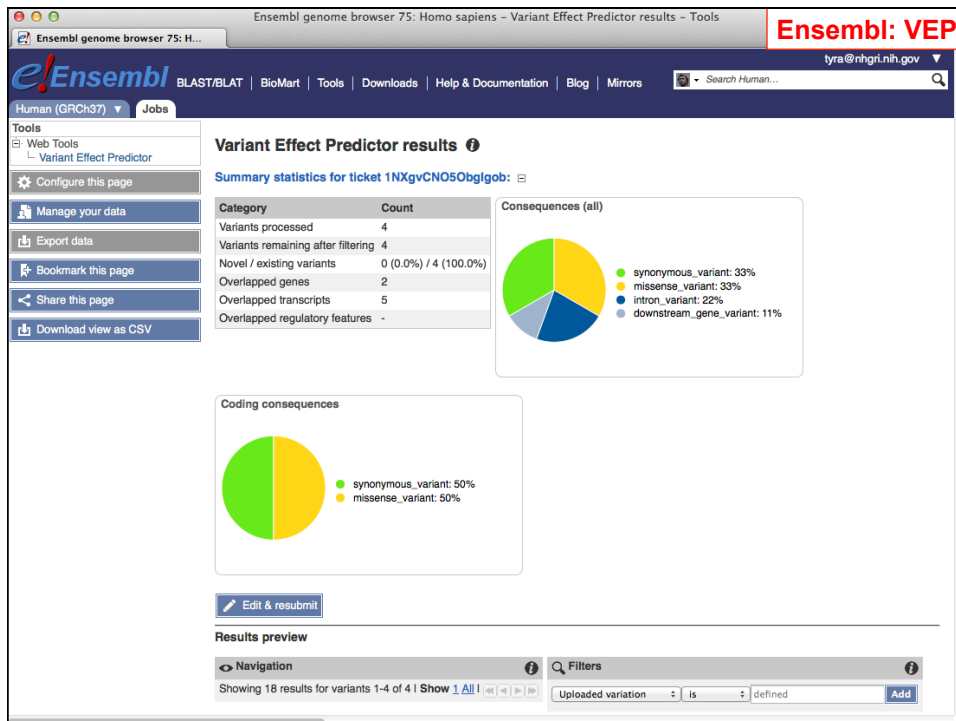
Or select previously uploaded file: --- Select file ---

Transcript database to use: Ensembl transcripts
 RefSeq and other transcripts

Output options

Identifiers and frequency data Additional identifiers for genes, transcripts and variants; frequency data

Extra options e.g. SIFT, PolyPhen and regulatory data



Uploaded variation	Location	Allele	Gene	Feature	Feature type	Consequence	cDNA position	CDS position	Protein position	Amino acids
rs144646998	8:39602389	C	ENSG00000104755	ENST00000521880	Transcript	missense_variant	2051	2009	670	P/R
rs144646998	8:39602389	C	ENSG00000104755	ENST00000265708	Transcript	missense_variant	2302	2198	733	P/R
rs144646998	8:39602389	C	ENSG00000104755	ENST00000347580	Transcript	missense_variant	2165	2141	714	P/R
rs144646998	8:39602389	C	ENSG00000104755	ENST00000379853	Transcript	missense_variant	1755	1730	577	P/R
rs145143599	8:39613262	G	ENSG00000104755	ENST00000521880	Transcript	Intron_variant	-	-	-	-
rs145143599	8:39613262	G	ENSG00000104755	ENST00000265708	Transcript	synonymous_variant	1886	1782	594	S
rs145143599	8:39613262	G	ENSG00000104755	ENST00000347580	Transcript	synonymous_variant	1749	1725	575	S
rs145143599	8:39613262	G	ENSG00000221018	ENST00000408091	Transcript	downstream_gene_variant	-	-	-	-
rs145143599	8:39613262	G	ENSG00000104755	ENST00000379853	Transcript	synonymous_variant	1339	1314	438	S
rs35935433	8:39613396	T	ENSG00000104755	ENST00000521880	Transcript	Intron_variant	-	-	-	-
rs35935433	8:39613396	T	ENSG00000104755	ENST00000265708	Transcript	missense_variant	1752	1648	550	V/I
rs35935433	8:39613396	T	ENSG00000104755	ENST00000347580	Transcript	missense_variant	1615	1591	531	V/I
rs35935433	8:39613396	T	ENSG00000221018	ENST00000408091	Transcript	downstream_gene_variant	-	-	-	-
rs35935433	8:39613396	T	ENSG00000104755	ENST00000379853	Transcript	Intron_variant	-	-	-	-

Amino acids	Codons	Existing variation	AA MAF	EA MAF	Symbol	SIFT	PolyPhen	GMAF	Biotype	AFR MAF	AMR MAF	ASN MAF	EUR MAF
P/R	CCT/CGT	rs144646998	-	-	ADAM2	0.03	0	C:0.0005	protein_coding	0.0020	0	0	0
P/R	CCT/CGT	rs144646998	-	-	ADAM2	0.03	0	C:0.0005	protein_coding	0.0020	0	0	0
P/R	CCT/CGT	rs144646998	-	-	ADAM2	0.02	0	C:0.0005	protein_coding	0.0020	0	0	0
P/R	CCT/CGT	rs144646998	-	-	ADAM2	0.02	0	C:0.0005	protein_coding	0.0020	0	0	0
-	-	rs145143599	0.0158874	0.000116279	ADAM2	-	-	G:0.0028	protein_coding	0.01	0.0028	0	0
S	TCT/TCC	rs145143599	0.0158874	0.000116279	ADAM2	-	-	G:0.0028	protein_coding	0.01	0.0028	0	0
S	TCT/TCC	rs145143599	0.0158874	0.000116279	ADAM2	-	-	G:0.0028	protein_coding	0.01	0.0028	0	0
-	-	rs145143599	0.0158874	0.000116279	AC136365.1	-	-	G:0.0028	miRNA	0.01	0.0028	0	0
S	TCT/TCC	rs145143599	0.0158874	0.000116279	ADAM2	-	-	G:0.0028	protein_coding	0.01	0.0028	0	0
-	-	rs35935433	0.000227066	0	ADAM2	-	-	-	protein_coding	-	-	-	-
V/I	GTA/ATA	rs35935433	0.000227066	0	ADAM2	0.03	0.004	-	protein_coding	-	-	-	-
V/I	GTA/ATA	rs35935433	0.000227066	0	ADAM2	0.06	0.047	-	protein_coding	-	-	-	-
-	-	rs35935433	0.000227066	0	ADAM2	-	-	-	protein_coding	-	-	-	-
-	-	rs35935433	0.000227066	0	ADAM2	-	-	-	protein_coding	-	-	-	-
L	TTA/CTA	rs3441791	-	-	ADAM2	-	-	-	protein_coding	-	-	-	-
L	TTA/CTA	rs3441791	-	-	ADAM2	-	-	-	protein_coding	-	-	-	-
L	TTA/CTA	rs3441791	-	-	ADAM2	-	-	-	protein_coding	-	-	-	-

Ensembl genome browser 75: Homo sapiens - Explore this variation - rs35935433

Ensembl Variation tab: Summary

Human (GRCh37) Location: 8:39,612,896-39,613,896 Variation: rs35935433

Variation displays

- Explore this variation
 - Genomic context
 - Genes and regulation (5)
 - Flanking sequence
 - Population genetics
 - Individual genotypes (34)
 - Linkage disequilibrium
 - Phenotype Data
 - Phylogenetic Context (6)
 - Citations
 - External Data
 - SNPedia
 - LOVD
- Configure this page
- Manage your data
- Export data
- Bookmark this page
- Share this page

rs35935433 SNP

Original source: Variants (including SNPs and indels) imported from dbSNP (release 138) | [View in dbSNP](#)

Alleles: C/T | Ambiguity code: Y

Location: Chromosome 8:39613396 (forward strand) | [View in location tab](#)

Most severe consequence: Missense variant | [See all predicted consequences \(Genes and regulation\)](#)

Evidence status:

HGVSN names: This variation has 7 HGVSN names - click the plus to show

Explore this variation

- Genomic context
- Genes and regulation
- Population genetics
- Individual genotypes
- Linkage disequilibrium
- Phenotype data
- Citations
- Phylogenetic context
- Flanking sequence

Using the website

- Video: [Browsing SNPs and CNVs in Ensembl](#)
- Video: [Clip: Genome Variation](#)
- Video: [BioMart: Variation IDs to HGNC Symbols](#)
- Exercise: [Genomes and SNPs in Malaria](#)

Analysing your data

Programmatic access

- Tutorial: [Accessing variation data with the Variation API](#)

Reference materials

- [Ensembl variation documentation portal](#)
- [Ensembl variation data description](#)
- [Variation Quick Reference card](#)

LOVD

- Configure this page
- Add your data
- Export data
- Bookmark this page
- Share this page
- Download view as CSV

Context

Genes (GENCODE...)

< ADAM2-003 protein coding

< ADAM2-004 protein coding

< ADAM2-001 protein coding

< ADAM2-002 protein coding

Sequence variant...

Ensembl Variation tab:
Genomic context

Splice acceptor variant	Stop gained	Frameshift variant
Missense variant	Splice region variant	Synonymous variant
Mature miRNA variant	Non coding exon variant	Intron variant

Share this page

Download view as RTF

BLAST this sequence

Key

Intronic	Mature miRNA	Missense
Synonymous	Splice acceptor	Splice region

• **Focus variant**

```

TCACCTGGCCAACTGTGAACTCCCTCTCTACTAAAAATACAAATTAGCGGGTG
TGGGTGACACCTGTAATCCAGCTACTCAAGAGCTGAGGCAGAGAACTACTGAAAC
CTGGTGGCAGAGGTGCAGTAGAGATCAGCAGCTGCACTTACCCAGGGCAGAGA
GTGCACCTACTCAGAAAGGGGAGGTCTAAATTTAGCAATTAAGGATATAAA
AGACTACCTACTTATTGAACCACAAAGTTCACCTTTTATCCAATCTTTGGCTG
TCTGCATGATCACGGCAATTCCACAGCAATGCAGAGATGTCCCTTATGTTGGCATAA
ATAATAGTGCCTTGGAAATTGTAATAAAAATTTACCTATATATACAAATTAATTTT
CCACTGCAGATTCATAATTTATCCAAATAAAAATTTATTAATTTAATTTAGAGA
AGTACAGATATATATTTTAAATAAGTAGATTTAAATTAAGTATTCAGGTTCCAG
CTAATACATCAATTTGTTTAAACACCAAAAAGAGCCTGCCAATTCCTCCATAA
TAACTGAACATTTTAGTAATTAATTAAGACTAAGGTTGGTCCAAACTAATAGTGCCTT
TTGCCACTTTTAAATGGCAAAAGCCCAATCCTTTGCACCACTAATATCCCTAG
TTAGCCTGACAACTTTTTTTGTGATTCCTTCACTTTCCTTTGAATCTGAAACAAATCAA
ACCAAGCTCAATGAATTA
                    
```

Ensembl Variation tab:
Flanking sequence

Ensembl

Location tab

<http://www.ensembl.org>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
 Division of Intramural Research

Ensembl genome browser 75: Homo sapiens - Region in detail - Chromosome 8: 39,613,346-39,613,446

Ensembl Location tab: Region in detail

click

Ensembl genome browser 75: Homo sapiens - Region in detail - Chromosome 8: 39,613,346-39,613,446

Ensembl Location tab: Configure page

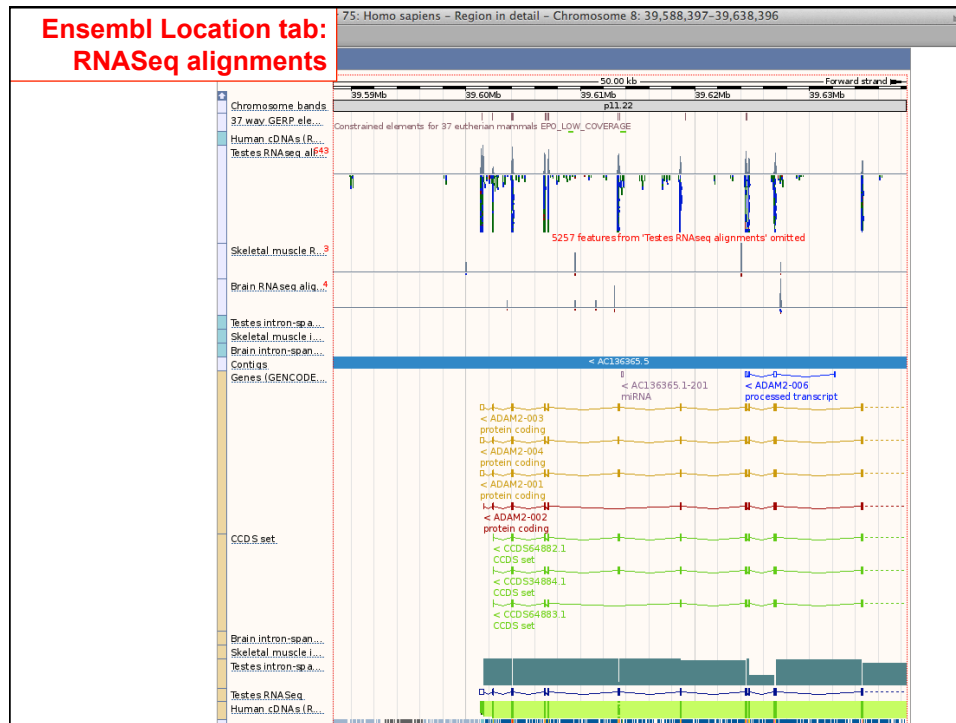
Configure Region Image | Configure Overview Image | Configure Chromosome Image | Personal Data

- Genes and transcripts (11/85)
 - Genes (2/6)
 - Prediction transcripts (0/2)
 - RNASeq models (9/77)
 - mRNA and protein alignments (1/14)
 - mRNA alignments (1/3)
 - EST alignments (0/1)
 - Protein alignments (0/6)
 - Protein features (0/6)
- ncRNA (0/1)
- Variation (0/91)
 - dbSNP (0/2)
 - 1000 Genomes & HapMap (0/13)
 - Phenotype and curated variants (0/17)
 - Individual genomes (0/14)
 - Arrays and other (0/13)
 - Failed variants (0/1)
 - Sequence variants (0/2)
 - Structural variants (0/26)
 - Recombination & Accessibility (0/3)
- Somatic mutations (0/5)
 - Somatic variants (0/3)
 - Somatic structural variants (0/2)
- Regulation (1/177)
 - Regulatory features (1/20)
 - Open chromatin & TFBS (0/14)
 - Histones & polymerases (0/13)
 - DNA Methylation (0/65)
 - Other regulatory regions (0/5)
- Comparative genomics (1/73)
 - Multiple alignments (0/4)
 - Conservation regions (1/5)
 - BLASTz/LASTz alignments (0/47)
 - Translated blat alignments (0/17)
- Oligo probes (0/37)
- Repeat regions (0/12)

Default style: Ensembl/Ensembl

Human BodyMap 2.0

Organ	Adipose	Adrenal	Blood	Brain	Breast	Colon	Heart	Kidney	Liver	Lung	Lymph	Merged (incl.Pooled)	Ovary	Pooled	Prostate	Skeletal muscle	Testes
Adipose	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Adrenal	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Blood	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Brain	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Breast	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Colon	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Heart	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Kidney	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Liver	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Lung	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Lymph	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Merged (incl.Pooled)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ovary	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Pooled	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Prostate	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Skeletal muscle	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Testes	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1



Ensembl

Gene tab

<http://www.ensembl.org>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Ensembl Location tab: Region in detail

Location: 8:39613346-39613446
 Gene: ADAM2

Chromosome bands: 37 way_GERP_ele...

Sequence: Constrained elements for 37 eutherian mammals EPO_LOW_COVERAGE

Contigs: < AC136355.5

Genes (GENCODE 19): < ADAM2-003 protein coding, < ADAM2-001 protein coding, < ADAM2-002 protein coding, < ADAM2-004 protein coding, < ADAM2-005 protein coding, < ADAM2-006 protein coding

CCDS.set: < CCDS-2672 Known protein coding, < CCDS-735 Known protein coding

Human cDNAs (RefSeq): < ADAM2-001

%GC: Prediction method: Ensembl/Havana merge Transcript

Gene Legend: pf, ct

Blocks show the locations of G/C base pairs

Ensembl Gene tab: Gene summary

Human (GRCh37) Location: 8:39,613,346-39,613,446 Gene: ADAM2 Transcript: ADAM2-001 Variation: rs393954

Gene-based displays: Summary, Splice variants (6), Transcript comparison, Supporting evidence, Sequence, Secondary Structure, External references, Regulation, Expression, Comparative Genomics, Genomic alignments, Gene tree (image), Gene tree (text), Gene tree (align), Gene gain/loss tree, Orthologues (54), Paralogue (6), Protein families (1), Phenotype, Genetic Variation, Variation table, Structural variation, External data, Personal annotation, ID History, Gene history

Gene: ADAM2 ENSG00000104755

Description: ADAM metallopeptidase domain 2 [Source:HGNC Symbol;Acc:198]

Location: Chromosome 8: 39,601,254-39,695,808 reverse strand.

INSDC coordinates: chromosome:GRCh37:CM000670.1:39601254:39695808:1

Transcripts: This gene has 6 transcripts (splice variants) Hide transcript table

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS	GENCODE basic
ADAM2-001	ENST00000265708	2672	ENSP00000265708	735	Protein coding	CCDS34884	Y
ADAM2-003	ENST00000347580	2535	ENSP00000343854	716	Protein coding	CCDS64882	Y
ADAM2-002	ENST00000521980	2125	ENSP00000429352	672	Protein coding	CCDS64883	Y
ADAM2-004	ENST00000373953	2125	ENSP00000369182	579	Protein coding	-	Y
ADAM2-005	ENST00000523181	728	No protein product	-	Processed transcript	-	Y
ADAM2-006	ENST00000520434	520	No protein product	-	Processed transcript	-	Y

Summary

Name: ADAM2 (HGNC Symbol)

Synonyms: CT15, FTNB, PH-30b, PH30 [To view all Ensembl genes linked to the name click here.]

CCDS: This gene is a member of the Human CCDS set: CCDS34884, CCDS64882, CCDS64883

RefSeq: Overlapping RefSeq Gene ID 2515 matches and has similar biotype of protein_coding

Ensembl version: ENSG00000104755.10

Gene type: Known protein coding

Prediction Method: Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.

Alternative genes: This gene corresponds to the following database identifiers:
 Havana gene: QTTTHUMG00000164041 (version 3)

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

Ensembl genome browser 75: Homo sapiens - Orthologues - Gene: ADAM2 (ENSG00000101411)

Ensembl Gene tab: Orthologues

Species set: Show details 1:1 1:many

Species set	1:1	1:many
Primates Humans and other primates	8	0
Rodents Rodents, rabbits and related species	7	0
Laurasiatheria Carnivores, ungulates and insectivores	14	0
Placental Mammals All placental mammals	34	0
Sauropsida Birds and Reptiles	2	5
Fish Ray-finned fishes	0	0
All All species, including invertebrates	39	13

Selected orthologues

Species	Type	dN/dS	Ensembl Identifier & gene name	Compare	Location	Target %id	Query %id
Anole lizard (<i>Anolis carolinensis</i>)	1-to-many	n/a	ENSACAG00000009283 Novel Ensembl prediction Uncharacterized protein [Source: UniProtKB/TrEMBL; acc: H9GF71]	Region Comparison Alignment (protein) Alignment (cDNA) Gene Tree (image)	GL343418.1:598118-623334:1	42	43
Anole lizard (<i>Anolis carolinensis</i>)	1-to-many	n/a	ENSACAG000000029425 Novel Ensembl prediction Uncharacterized protein [Source: UniProtKB/TrEMBL; acc: B4GDH5]	Region Comparison Alignment (protein) Alignment (cDNA) Gene Tree (image)	GL343418.1:650240-673477:1	38	29
Chicken (<i>Gallus gallus</i>)	1-to-many	n/a	ENSGALG00000003444 Novel Ensembl prediction Uncharacterized protein [Source: UniProtKB/TrEMBL; acc: F1NP23]	Region Comparison Alignment (protein) Alignment (cDNA) Gene Tree (image)	22:2443241-2448175:1	38	36

click

Ensembl genome browser 75: Homo sapiens - Variation image - Gene: ADAM2 (ENSG00000101411)

Ensembl Gene tab: Variation Image

CSMIC Variations

Genes (GENCODE):

ENST00000265708
ADAM2-001

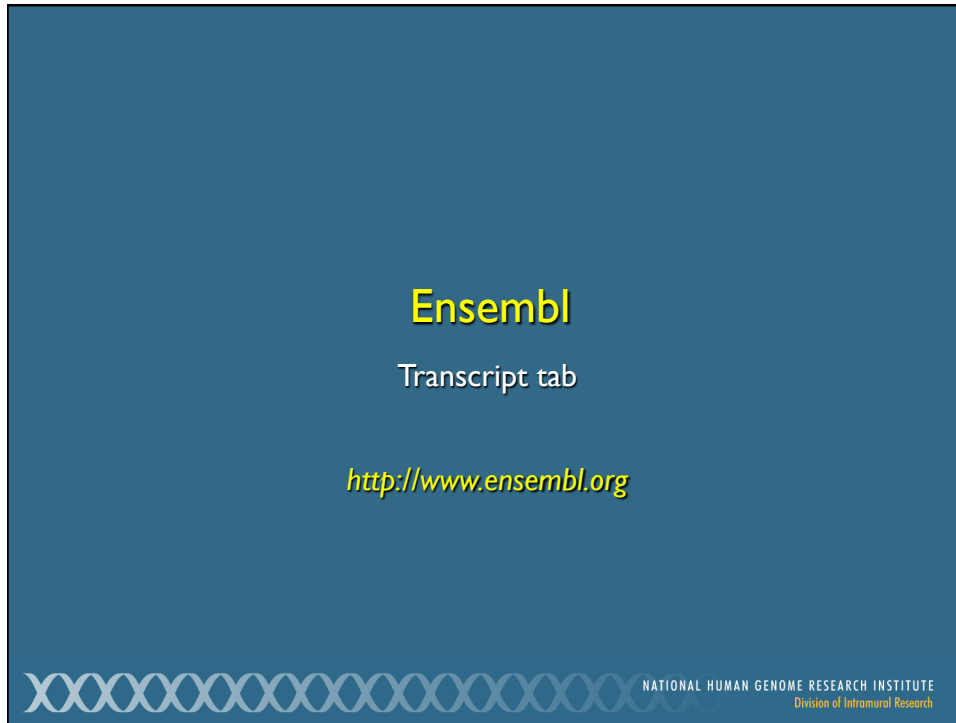
Variation Legend:

- Splice donor variant
- Frameshift variant
- Splice region variant
- 5 prime UTR variant
- Intron variant
- Regulatory region variant
- Splice acceptor variant
- Inframe deletion
- Synonymous variant
- 3 prime UTR variant
- Upstream gene variant
- Intergenic variant
- Stop gained
- Missense variant
- Mature miRNA variant
- Non coding exon variant
- Downstream gene variant

PROSITE profiles

Pfam domain

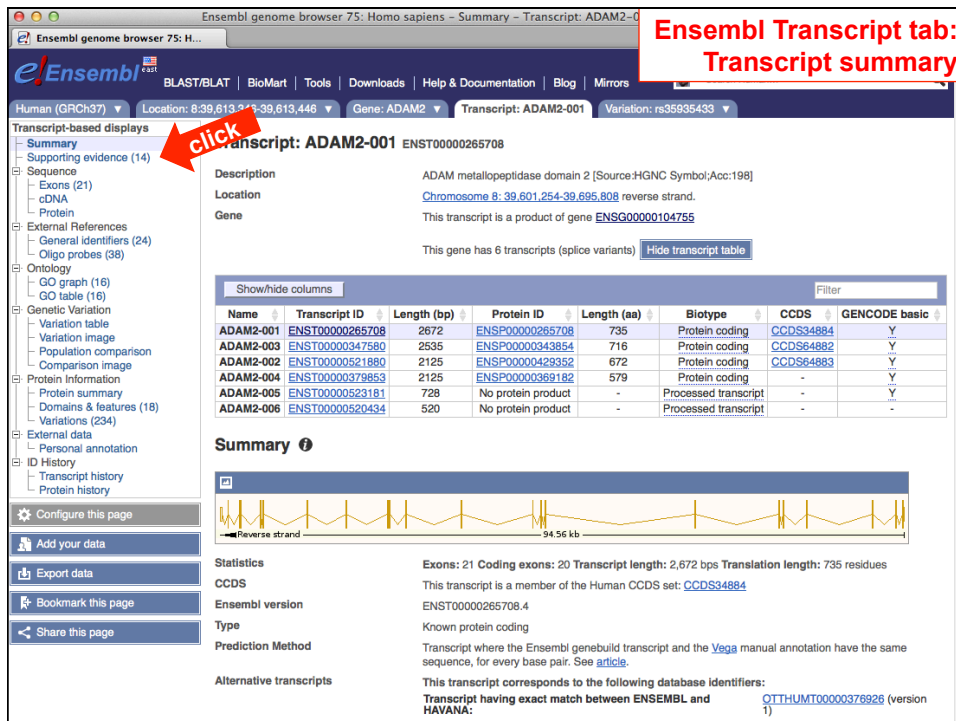
Prints domain



Ensembl
 Transcript tab

<http://www.ensembl.org>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
 Division of Intramural Research



Ensembl Transcript tab:
 Transcript summary

click

Transcript: ADAM2-001 ENST00000265708

Description ADAM metallopeptidase domain 2 [Source:HGNC Symbol;Acc:198]
Location [Chromosome 8: 39,601,254-39,695,808 reverse strand.](#)
Gene This transcript is a product of gene [ENSG000001104755](#)

This gene has 6 transcripts (splice variants) [Hide transcript table](#)

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS	GENCODE basic
ADAM2-001	ENST00000265708	2672	ENSP00000265708	735	Protein coding	CCDS34884	Y
ADAM2-003	ENST00000347580	2535	ENSP00000343854	716	Protein coding	CCDS64882	Y
ADAM2-002	ENST00000521880	2125	ENSP00000428352	672	Protein coding	CCDS64883	Y
ADAM2-004	ENST00000379853	2125	ENSP00000369182	579	Protein coding	-	Y
ADAM2-005	ENST00000523181	728	No protein product	-	Processed transcript	-	Y
ADAM2-006	ENST00000520434	520	No protein product	-	Processed transcript	-	-

Summary

Statistics
 Exons: 21 Coding exons: 20 Transcript length: 2,672 bps Translation length: 735 residues
 CCDS This transcript is a member of the Human CCDS set: [CCDS34884](#)
 Ensembl version ENST00000265708.4
 Type Known protein coding
 Prediction Method Transcript where the Ensembl genebuild transcript and the [Vega](#) manual annotation have the same sequence, for every base pair. See [article](#).
 Alternative transcripts This transcript corresponds to the following database identifiers:
 Transcript having exact match between ENSEMBL and HAVANA: [OTTNUMT00000376926](#) (version 1)

Ensembl genome browser 75: Homo sapiens - Supporting evidence - Transcript: ADAM2-001

Ensembl Transcript tab: Supporting evidence

Human (GRCh37) Location: 8:39,613,346-39,613,446 Gene: ADAM2 Transcript: ADAM2-001 Variation: rs35935433

Transcript: ADAM2-001 ENST00000265708

Description ADAM metalloproteinase domain 2 [Source:HGNC Symbol;Acc:198]
 Location [Chromosome 8: 39,601,254-39,695,808](#) reverse strand.
 Gene This transcript is a product of gene [ENSG00000104755](#)
 This gene has 6 transcripts (splice variants) [Show transcript table](#)

Supporting evidence

Legend: protein evidence (yellow), EST evidence (purple), cDNA evidence (green)

- non-canonical splice site
- evidence start / ends within exon / CDS
- evidence extends beyond exon / CDS
- part of evidence duplicated in transcript structure
- part of evidence missing from transcript structure
- evidence extends beyond the end of the transcript

[Click here for a summary of the evidence that supports all the transcripts of this gene.](#)

Ensembl genome browser 75: Homo sapiens - Protein sequence - Transcript: ADAM2-001

Ensembl Transcript tab: Protein sequence

Human (GRCh37) Location: 8:39,613,346-39,613,446 Gene: ADAM2 Transcript: ADAM2-001 Variation: rs35935433

Transcript: ADAM2-001 ENST00000265708

Description ADAM metalloproteinase domain 2 [Source:HGNC Symbol;Acc:198]
 Location [Chromosome 8: 39,601,254-39,695,808](#) reverse strand.
 Gene This transcript is a product of gene [ENSG00000104755](#)
 This gene has 6 transcripts (splice variants) [Show transcript table](#)

Protein sequence

Key: Exons Alternating exons Residue overlap splice site

```

MWRVFLFLLSGLGRLRMDSNFDSLVPQIVPEKIRSIIEKIGESQASYKIVIEGRPYTVNL
MQKFLPHNFRVYSYSGTGIMKPLDQDFNCFYQYIEGPKSVVMVSTCIGLRGVLP
ENVSYGIEPLESSVGFHVIYQVKKKADVSLYNEKDIESRDLSPFLQSVFQQDFARYI
EMGVIVKQLNIMSGSTTVVQKVPQLIGLNIFVSNPITLISLEIMIDENKIAIT
GEANELLHFLRWRKTSYLVLRPHDVALLVYREKSNVYVGFQGMCDANVAGCVLHPR
TISLESLAVILQALLSMGIYDDINKQCQSGVAVIMNPEAIFHSGVKIFNSCFEFA
HFSKQKSQLINQFRLLDPFFKQAVCONAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCQCNLFMSKERMCRPSFECDLPEYCNSSASCPENHYVQVGHPCGLN
QWLCIDVCMGSGKCTDFFGRVEVFGPSECHSHLNSKIDVSGNCSISDSYFQCADNL
QCKLICKYVYKPLQIPRATLIYANISGHLIAVEFASDHADSQMMKIKDGTSCGSKNV
CRNQRCVSSYLGYDQTDKCDNRGVCNKKKHCASLYPPDCSVQSDLPFGVSDSGN
PPPVAIPARLERRVYIENHSKPMRWFPLFIFDFIIFCVLAIMVKVNFQRKKWRTE
YSSDQPESESEPKG
    
```

Ensembl release 75 - February 2014 [View in archive site](#)

[About Ensembl](#) | [Privacy Policy](#) | [Contact Us](#)

click

The screenshot shows the Ensembl genome browser interface. The browser title is "Ensembl genome browser 75: Homo sapiens - Protein sequence - Transcript: ADAM2-001 (ENST000...". The main content area is titled "View in archive site" and contains a list of available Ensembl archives. The list includes:

- Ensembl 74: Dec 2013 (GRCh37) - patched/updated gene set Sep 2013
- Ensembl 73: Sep 2013 (GRCh37) - patched/updated gene set Jun 2013
- Ensembl 72: Jun 2013 (GRCh37) - patched/updated gene set Apr 2013
- Ensembl 71: Apr 2013 (GRCh37) - patched/updated gene set Feb 2013
- Ensembl 70: Jan 2013 (GRCh37)
- Ensembl 69: Oct 2012 (GRCh37) - patched/updated gene set Oct 2012
- Ensembl 68: Jul 2012 (GRCh37) - patched/updated gene set May 2012
- Ensembl 67: May 2012 (GRCh37)
- Ensembl 66: Feb 2012 (GRCh37) - patched/updated gene set Feb 2012
- Ensembl 65: Dec 2011 (GRCh37) - patched/updated gene set Dec 2011
- Ensembl 64: Sep 2011 (GRCh37) - patched/updated gene set Sep 2011
- Ensembl 63: Jun 2011 (GRCh37) - gene set updated Apr 2011
- Ensembl 62: Apr 2011 (GRCh37) - gene set updated Apr 2011
- Ensembl 59: Aug 2010 (GRCh37)
- Ensembl 54: May 2009 (NCBI36)

Below the list is a link: [More information about the Ensembl archives](#). The interface also includes a search bar, navigation tabs (Human, Transcript, Variation), and a sidebar with various tools and options.

Ensembl

Find a chicken homolog of a human protein

<http://www.ensembl.org>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Query	Subject	Chromosome	Scaffold	Contig	Stats	Sort By
off	_off_	_off_	_off_	_off_	_off_	
Name	Name	Name	Name	Name	Name	
Start	Start	Start	Start	Start	E-val	>Contig
						<Score
						>Score
[A] [S] [G] [C]	6 662 +	Chr:15	6235722	6237707 +	1365 8.4e-134 35.33 685	
[A] [S] [G] [C]	2 672 +	Chr:15	6239085	6241091 +	1352 5.9e-131 34.01 691	
[A] [S] [G] [C]	4 683 +	Chr:5	26725477	26727474 +	1312 4.6e-126 31.82 682	
[A] [S] [G] [C]	138 647 +	Chr:22	2444846	2446357 +	300 1.0e-23 24.50 551	
[A] [S] [G] [C]	370 644 +	Chr:17				
[A] [S] [G] [C]	330 651 +	Chr:17				
[A] [S] [G] [C]	400 511 +	Chr:22				
[A] [S] [G] [C]	386 570 +	Chr:22				
[A] [S] [G] [C]	330 673 +	Chr:26				
[A] [S] [G] [C]	438 627 +	Chr:22				
[A] [S] [G] [C]	410 674 +	Chr:4				
[A] [S] [G] [C]	232 659 +	Chr:1				
[A] [S] [G] [C]	445 505 +	Chr:22				
[A] [S] [G] [C]	330 687 +	Chr:25				
[A] [S] [G] [C]	328 668 +	No data				
[A] [S] [G] [C]	378 531 +	Chr:8				
[A] [S] [G] [C]	438 502 +	Chr:6				
[A] [S] [G] [C]	387 652 +	No data				
[A] [S] [G] [C]	438 651 +	Chr:13				
[A] [S] [G] [C]	209 413 +	Chr:22				
[A] [S] [G] [C]	374 603 +	Chr:22				
[A] [S] [G] [C]	407 652 +	Chr:6				
[A] [S] [G] [C]	444 501 +	Chr:22				
[A] [S] [G] [C]	339 443 +	Chr:22				

BioMart
Cross-reference data from different sources

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

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

BioMart
Get genomic coordinates, gene name, and RefSeq accessions for ENSEMBL gene identifiers

Step 1: Select Dataset

Please restrict your query using criteria below

Step 2: Select Filters (input)

The screenshot shows the Ensembl BioMart interface. The top navigation bar includes 'Ensembl', 'BLAST/BLAT', 'BioMart', 'Tools', 'Downloads', 'Help & Documentation', and 'Blog'. Below this is a search bar and a 'Search' button. The main interface is divided into several sections. On the left, there are tabs for 'New', 'Count', and 'Results', along with buttons for 'URL', 'XML', 'Perl', and 'Help'. The 'Dataset' section on the left shows 'Danio rerio genes (Zv9)' selected. The main area is titled 'Please restrict your query using criteria below'. It has sections for 'REGION' and 'GENE'. Under 'GENE', there are options for 'Limit to genes ...' and 'with ArrayExpress ID(s)'. A red box highlights the 'ID list limit [Max 500 advised]' checkbox, which is checked, and a list of Ensembl Gene IDs: ENSDARG00000086764, ENSDARG00000062831, ENSDARG00000069528, ENSDARG00000075385, and ENSDARG00000019658. Below this, there are options for 'Transcript count >=', 'Gene type' (with a dropdown menu showing 'antisense', 'IC_C_pseudogene', 'IC_L_pseudogene', 'IC_V_pseudogene', and 'lincRNA'), and 'Source (gene)' (with a dropdown menu showing 'ensembl').

BioMart
 Get genomic coordinates, gene name, and RefSeq accessions for ENSEMBL gene identifiers

Dataset: Danio rerio genes (Zv9)

Filters: Ensembl Gene ID(s) [e.g. ENSG00000139618]; [ID-list specified]

Attributes: Ensembl Gene ID, Ensembl Transcript ID, Chromosome Name, Gene Start (bp), Gene End (bp), Associated Gene Name, RefSeq mRNA [e.g. NM_001195597], RefSeq mRNA predicted [e.g. XM_001125684]

Please select columns to be included in the output and hit 'Go'

Features: Features, Structures, Transcript Event

Homologs: Homologs, Variation, Sequences

GENE: Ensembl

- Ensembl Gene ID
- Ensembl Transcript ID
- Ensembl Protein ID
- Ensembl Exon ID
- Description
- Chromosome Name
- Gene Start (bp)
- Gene End (bp)
- Strand
- Band
- Associated Gene Name
- Associated Transcript Name
- Associated Gene DB
- Associated Transcript DB
- Transcript count
- % GC content
- Gene Biotype
- Transcript Biotype
- Source (gene)
- Source (transcript)

External References (max 3)

- ArrayExpress
- ChEMBL ID(s)
- Clone based Ensembl gene name
- Clone based Ensembl transcript name
- Clone based VEGA gene name
- Clone based VEGA transcript name
- EMBL (Genbank) ID
- EntrezGene ID
- VEGA gene ID(s) (OTTG)
- VEGA transcript ID(s) (OTTT)
- VEGA protein ID(s) (OTTP)
- HGNC ID(s)
- HGNC symbol
- HGNC transcript name
- MEROPS ID
- PDB ID
- miRBase Accession(s)
- miRBase ID(s)
- miRBase gene name
- miRBase transcript name
- Protein (Genbank) ID
- RefSeq mRNA [e.g. NM_001195597]
- RefSeq mRNA predicted [e.g. XM_001125684]
- RefSeq ncRNA [e.g. NR_002834]
- RefSeq ncRNA predicted [e.g. XR_108264]
- RefSeq Protein ID [e.g. NP_001005353]
- RefSeq Predicted Protein ID [e.g. XP_001720922]
- Rfam ID
- Rfam gene
- Rfam tran
- Unigene I
- UniProt/TrEMBL Accession
- UniProt/SwissProt ID
- UniProt/SwissProt Accession
- UniProt Gene Name
- Uniprot Genename Transcript Name
- UniParc
- WikiGene Name
- WikiGene ID
- WikiGene Description
- ZFIN ID
- ZFIN symbol
- ZFIN transcript name

Step 3: Select Attributes (output)

BioMart
 Get genomic coordinates, gene name, and RefSeq accessions for ENSEMBL gene identifiers

Dataset: Danio rerio genes (Zv9)

Filters: Ensembl Gene ID(s) [e.g. ENSG00000139618]; [ID-list specified]

Attributes: Ensembl Gene ID, Ensembl Transcript ID, Chromosome Name, Gene Start (bp), Gene End (bp), Associated Gene Name, RefSeq mRNA [e.g. NM_001195597], RefSeq mRNA predicted [e.g. XM_001125684]

Export all results to: File, TSV, Unique results only

Email notification to: []

View: 50 rows as HTML, Unique results only

Ensembl Gene ID	Ensembl Transcript ID	Chromosome Name	Gene Start (bp)	Gene End (bp)	Associated Gene Name	RefSeq mRNA [e.g. NM_001195597]	RefSeq mRNA predicted [e.g. XM_001125684]
ENSDARG00000000906	ENSDART000000052660	16	23018783	23062136	skap2	NM_200628	XM_005157963
ENSDARG00000000906	ENSDART00000137344	16	23018783	23062136	skap2		
ENSDARG00000002006	ENSDART00000021596	16	20493224	20528393	rxrbb	NM_131238	
ENSDARG00000002006	ENSDART00000147944	16	20493224	20528393	rxrbb		
ENSDARG00000002006	ENSDART00000128914	16	20493224	20528393	rxrbb		XM_005157945
ENSDARG00000002507	ENSDART00000139859	16	16045949	16118555	itga10		XM_005157918
ENSDARG00000002507	ENSDART00000139859	16	16045949	16118555	itga10		XM_003200156
ENSDARG00000002507	ENSDART00000011224	16	16045949	16118555	itga10		
ENSDARG00000004358	ENSDART00000012673	16	13772550	13799769	grnb3a	NM_001002437	XM_005173480
ENSDARG00000004561	ENSDART00000142610	16	14772197	14861170	prkcg		XM_001921680
ENSDARG00000004561	ENSDART00000103886	16	14772197	14861170	prkcg		
ENSDARG00000004806	ENSDART00000121998	16	15611220	15622320	grwd1	NM_001003509	
ENSDARG00000005762	ENSDART00000139611	16	16979935	17345861	col14a1a		
ENSDARG00000005762	ENSDART00000137912	16	16979935	17345861	col14a1a		
ENSDARG00000005762	ENSDART00000134087	16	16979935	17345861	col14a1a		
ENSDARG00000005762	ENSDART00000027982	16	16979935	17345861	col14a1a		
ENSDARG00000006983	ENSDART00000148426	16	1357323	1386898	cellf3b		XM_002664766
ENSDARG00000006983	ENSDART00000148426	16	1357323	1386898	cellf3b		XM_005158471
ENSDARG00000006983	ENSDART00000024206	16	1357323	1386898	cellf3b		
ENSDARG00000007959	ENSDART00000137902	16	22956445	22973946	hibadhb		
ENSDARG00000007959	ENSDART00000009429	16	22956445	22973946	hibadhb	NM_201160	
ENSDARG00000007959	ENSDART00000132407	16	22956445	22973946	hibadhb		
ENSDARG00000007959	ENSDART00000131452	16	22956445	22973946	hibadhb		
ENSDARG00000009023	ENSDART00000146436	16	22143616	22239485	ankrd28b		XM_684152
ENSDARG00000009023	ENSDART00000027020	16	22143616	22239485	ankrd28b		
ENSDARG00000001371	ENSDART000000007842	16	14545332	14561307	isoc2	NM_001079953	
ENSDARG00000001371	ENSDART00000146997	16	14545332	14561307	isoc2		
ENSDARG000000018787	ENSDART000000015986	16	26521948	26537442	efna1b		
ENSDARG000000018787	ENSDART00000138279	16	26521948	26537442	efna1b		
ENSDARG000000019658	ENSDART000000141032	16	12823290	12882007	pou2f2b		

BioMart:
 Get predicted human orthologs for ENSEMBL gene identifiers

Please select columns to be included in the output and hit

Features Homologs
 Structures Variation
 Transcript Event Sequences

GENE:

ORTHOLOGS (Max select 6 orthologs):

Alpaca Orthologs

Alpaca Ensembl Gene ID Ancestor
 Canonical Protein or Transcript ID Orthology confidence [0 low, 1 high]
 Alpaca Ensembl Protein ID % Identity with respect to query gene
 Alpaca Chromosome Name % Identity with respect to Alpaca gene
 Alpaca Chromosome Start (bp) dN

Human Orthologs


Human Ensembl Gene ID Ancestor
 Canonical Protein or Transcript ID Orthology confidence [0 low, 1 high]
 Human Ensembl Protein ID % Identity with respect to query gene
 Human Chromosome Name % Identity with respect to Human gene
 Human Chromosome Start (bp) dN
 Human Chromosome End (bp) dS
 Homology Type

Ensembl Gene ID	Ensembl Transcript ID	Human Ensembl Gene ID	Human Ensembl Protein ID	% Identity with respect to query gene
ENSDARG00000000906	ENSDDART00000052660	ENSG00000005020	ENSP00000005587	59
ENSDARG00000000906	ENSDDART00000137344	ENSG00000005020	ENSP00000005587	59
ENSDARG00000002006	ENSDDART00000021596	ENSG000000204231	ENSP000000363817	70
ENSDARG00000002006	ENSDDART00000147844	ENSG000000204231	ENSP000000363817	70
ENSDARG00000002006	ENSDDART00000128914	ENSG000000204231	ENSP000000363817	70
ENSDARG00000002507	ENSDDART00000138859	ENSG00000143127	ENSP000000358310	54
ENSDARG00000002507	ENSDDART00000011224	ENSG00000143127	ENSP000000358310	54
ENSDARG00000004358	ENSDDART00000012673	ENSG00000069966	ENSP000000261837	49
ENSDARG00000004561	ENSDDART00000142610	ENSG00000126583	ENSP000000263431	69
ENSDARG00000004561	ENSDDART00000103886	ENSG00000126583	ENSP000000263431	69
ENSDARG00000004806	ENSDDART00000121998	ENSG00000105447	ENSP000000253237	59
ENSDARG00000005762	ENSDDART00000138611	ENSG00000187955	ENSP000000297848	60
ENSDARG00000005762	ENSDDART00000137912	ENSG00000187955	ENSP000000297848	60
ENSDARG00000005762	ENSDDART00000134087	ENSG00000187955	ENSP000000297848	60
ENSDARG00000005762	ENSDDART00000027982	ENSG00000187955	ENSP000000297848	60
ENSDARG00000006983	ENSDDART00000148426	ENSG00000159409	ENSP000000290583	77
ENSDARG00000006983	ENSDDART00000024206	ENSG00000159409	ENSP000000290583	77
ENSDARG00000007959	ENSDDART00000137902	ENSG00000106049	ENSP000000265395	76
ENSDARG00000007959	ENSDDART00000006429	ENSG00000106049	ENSP000000265395	76

Galaxy

knownGene coding sequence lengths

<https://usegalaxy.org/>


 NATIONAL HUMAN GENOME RESEARCH INSTITUTE
 Division of Intramural Research

Galaxy:
Step 1: Download transcript data from UCSC

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the [OpenHelix Table Browser tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal genome: Human assembly: Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Predictions track: UCSC Genes

table: knownGene describe table schema

region: genome ENCODE Pilot regions position chr21:33,031,597-33,041,570

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

correlation: create

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

chr1	11873	14409	uc001aaa.3	0	+	11873	11873	0	3
chr1	11873	14409	uc010nrx.1	0	+	11873	11873	0	3
chr1	11873	14409	uc010nxq.1	0	+	12189	13639	0	3
chr1	14361	16765	uc009vis.3	0	-	14361	14361	0	4
chr1	16857	17751	uc009vjc.1	0	-	16857	16857	0	2
chr1	15795	18061	uc009vjd.2	0	-	15795	15795	0	5
chr1	14361	19759	uc009vit.3	0	-	14361	14361	0	9
chr1	14361	19759	uc009vlu.3	0	-	14361	14361	0	10

Galaxy:
Step 2: Extract coding exons

Gene BED To Exon/Intron/Codon BED (version 1.0.0)

Extract: Coding Exons only

from: 1: UCSC Main on Human: knownGene (genome)

Execute

This tool works only on a BED file that contains at least 12 fields (see [Example](#) and [About formats](#) below). The output will be empty if applied to a BED file with 3 or 6 fields.

What it does

BED format can be used to represent a single gene in just one line, which contains the information about exons, coding sequence location (CDS), and positions of untranslated regions (UTRs). This tool *unpacks* this information by converting a single line describing a gene into a collection of lines representing individual exons, introns, UTRs, etc.

chr1	12189	12227	uc010nxq.1	0	+
chr1	12594	12721	uc010nxq.1	0	+
chr1	13402	13639	uc010nxq.1	0	+
chr1	69090	70008	uc001aal.1	0	+
chr1	138529	139696	uc021oeg.2	0	-
chr1	139789	139792	uc021oeg.2	0	-
chr1	324514	324686	uc021oeh.1	0	+
chr1	324718	325124	uc021oeh.1	0	+
chr1	325382	325605	uc021oeh.1	0	+
chr1	324342	324345	uc009vjk.2	0	+
chr1	324438	325605	uc009vjk.2	0	+
chr1	327745	328213	uc021oei.1	0	+

Galaxy:
Step 3: Calculate length of each coding exon

Compute (version 1.1.0)

Add expression:

as a new column to:

Dataset missing? See TIP below

Round result?:

TIP: If your data is not TAB delimited, use [Text Manipulation->Convert](#)

What it does
 This tool computes an expression for every row of a dataset and appends the result as a new column (field).

chr1	12189	12227	uc010nxq.1	0	38.0
chr1	12594	12721	uc010nxq.1	0	127.0
chr1	13402	13639	uc010nxq.1	0	237.0
chr1	69090	70008	uc001aal.1	0	918.0
chr1	138529	139696	uc021oeg.2	0	1167.0
chr1	139789	139792	uc021oeg.2	0	3.0
chr1	324514	324686	uc021oeh.1	0	172.0
chr1	324718	325124	uc021oeh.1	0	406.0
chr1	325382	325605	uc021oeh.1	0	223.0

Galaxy:
Step 4: Group coding exon lengths by transcript

Group (version 2.0.0)

Select data:

Dataset missing? See TIP below.

Group by column:

Ignore case while grouping?:

Operations

Operation 1

Type:

On column:

Round result to nearest integer?:

Group data by a column and perform aggregate operation on other columns.

uc001aal.1	918
uc001aa.3	1170
uc001abe.4	624
uc001abv.1	429
uc001abw.1	2046
uc001abx.2	1998
uc001aby.4	1656
uc001abz.4	2250

Galaxy:
 Create Workflow and run on other datasets

Running workflow "Workflow constructed from history 'knownGene coding sequence lengths'"

Step 1: Input dataset
 Input Dataset
 1: UCSC Main on Human: knownGene (genome)

Step 2: Gene BED To Exon/Intron/Codon BED (version 1.0.0)

Step 3: Compute (version 1.1.0)

Step 4: Group (version 2.0.0)

Send results to a new history
 Run workflow

Galaxy NGS tools

- Quality control and manipulation
 - FASTQC
- Mapping
 - Bowtie
- Peak Calling
 - MACS
- RNA-seq
 - TopHat
 - CuffLinks

NGS TOOLBOX BETA
[Phenotype Association](#)
[NGS: QC and manipulation](#)
[NGS: Mapping](#)
[NGS: SAM Tools](#)
[NGS: GATK Tools \(beta\)](#)
[NGS: Peak Calling](#)
[NGS: RNA-seq](#)
[NGS: Picard \(beta\)](#)
[NGS: Variant Analysis](#)
[snpEff](#)
[BEDTools](#)
[EMBOSS](#)

Additional resources

- UCSC Genome Browser User Guide
<http://genome.ucsc.edu/goldenPath/help/>
- Ensembl Tutorials and Worked Examples
<http://www.ensembl.org/info/website/tutorials/>
- Galaxy Support
<https://wiki.galaxyproject.org/Support/>



Current Protocols in Bioinformatics

<p>The UCSC Genome Browser</p>	<p>UNIT 1.4</p>
<p>Donna Karolchik,¹ Angie S. Hinrichs,¹ and W. James Kent¹</p> <p>¹Center for Biomolecular Science and Engineering, University of California Santa Cruz, California</p>	<p>Using Galaxy to Perform Large-Scale Interactive Data Analyses</p>
<p>ABSTRACT</p> <p>The University of California Santa Cruz (UCSC) Genome Browser based tool for quickly displaying a requested portion of a genome accompanied by a series of aligned annotation "tracks." The annotation tracks are provided by the UCSC Genome Bioinformatics Group and external collaborators and include gene models, mRNA and expressed sequence tag alignments, simple nucleotide repeat content, CpG islands, and other genomic features. The browser also displays expression and regulatory data, phenotype and variation data, and phylogenetic comparative genomics data. All information relevant to a region of the genome is displayed in a single window, facilitating biological analysis and interpretation. The underlying data can be viewed, downloaded, and analyzed using another Web-based application, the UCSC Table Browser. Users can also</p>	<p>Jennifer Hillman-Jackson,¹ Dave Clements,² Daniel Blankenberg,¹ James Taylor,² Anton Nekrutenko,¹ and Galaxy Team^{1,2}</p> <p>¹Penn State University, University Park, Pennsylvania ²Emory University, Atlanta, Georgia</p> <p>ABSTRACT</p> <p>Innovations in biomedical research technologies continue to provide experimental biologists with novel and increasingly large genomic and high-throughput data resources to be analyzed. As creating and obtaining data has become easier, the key decision faced by many researchers is a practical one: where and how should an analysis be performed? The Galaxy project was developed to address this question. Galaxy is a Web-based platform for data analysis and use is riddled with complexities outside of the traditional command-line environment. The authors believe that Galaxy provides a powerful means for data analysis in an intuitive Web application, integrating bioinformatics tools previously only available to command-line environments. We will demonstrate through examples how Galaxy specifically brings together (1) bioinformatics tools, for example, UCSC's Eukaryote and Genes tracks, (2) command-line tools (wrapped Unix functions, format converters), and (3) third-party analysis tools. <i>Curr. Protoc. Bioinform.</i> 30:1.15.1-1.15.48. © 2010 by John Wiley & Sons, Inc.</p>
<p>Using the Ensembl Genome Server to Browse Genomic Sequence Data</p> <p>Xosé M. Fernández-Suárez¹ and Michael K. Schuster¹</p> <p>¹EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom</p> <p>ABSTRACT</p> <p>The Ensembl project provides a comprehensive source of automatic annotation of the human genome sequence, as well as other species of biomedical interest, with confirmed gene predictions that have been integrated with external data sources. This unit describes how to use the Ensembl genome browser (http://www.ensembl.org), the public interface of the project. It describes how to find a gene or protein of interest, how to get additional information and external links, and how to use the comparative genomics tools. <i>Curr. Protoc. Bioinform.</i> 30:1.15.1-1.15.48. © 2010 by John Wiley & Sons, Inc.</p>	<p>UNIT 1.15</p> <p>bioinformatics • genomic alignments •</p>
<p>Access from NIH at http://onlinelibrary.wiley.com/book/10.1002/0471250953</p>	