

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\_GL000250y2\_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

### 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 [Neandertal](#) 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.

**Human (Homo sapiens) Genome Browser Gateway**

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

group: Mammal genome: Human assembly: Feb. 2009 (GRCh37/hg19) position: chr8:39601255-39695808 search term: adam2

ADAM2 (Homo sapiens ADAM metalloproteinase domain 2 (ADAM2), transcript variant 1, mRNA)  
 ADAM20 (Homo sapiens ADAM metalloproteinase domain 20 (ADAM20), mRNA)  
 ADAM20P1 (Homo sapiens ADAM metalloproteinase domain 20 pseudogene 1 (ADAM20P1), non-coding RNA)  
 ADAM21 (Homo sapiens ADAM metalloproteinase domain 21 (ADAM21), mRNA)  
 ADAM21P1 (Homo sapiens ADAM metalloproteinase domain 21 pseudogene 1 (ADAM21P1), non-coding RNA)  
 ADAM22 (Homo sapiens ADAM metalloproteinase domain 22 (ADAM22), transcript variant 1, mRNA)  
 ADAM23 (Homo sapiens ADAM metalloproteinase domain 23 (ADAM23), mRNA)  
 ADAM28 (Homo sapiens ADAM metalloproteinase domain 28 (ADAM28), transcript variant 1, mRNA)  
 ADAM29 (Homo sapiens ADAM metalloproteinase domain 29 (ADAM29), transcript variant 7, mRNA)

**Human Genome Browser - hg19 assembly (sequences)**

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI [Assembly database](#).

**Sample position queries**

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

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

UCSC  
 Homo sapiens  
 (Graphic courtesy of CBSI)

**UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly**

chr8:39,601,255-39,695,808 94,554 bp. enter position, gene symbol or search terms go [New! On-site workshops available.](#)

click

exon intron

move start < 2.0 > move end < 2.0 >

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

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

TFN Factor ChIP

100 Vert. Cons

Multiz 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 pack

RefSeq Genes dense hide hide dense squish pack

AceView Genes hide

CCDS hide

Ensembl Genes hide

EvoFold hide

Exoniphy hide

Geneid Genes hide

Genscan Genes hide

H-Inv 7.0 hide

IKMC Genes Mapped hide

click



Genomic Sequence Near Gene

UCSC RefSeq Gene details

Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

**Get Genomic Sequence Near Gene**

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

**Sequence Retrieval Region Options:**

- Promoter/Upstream by  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 
  - 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**

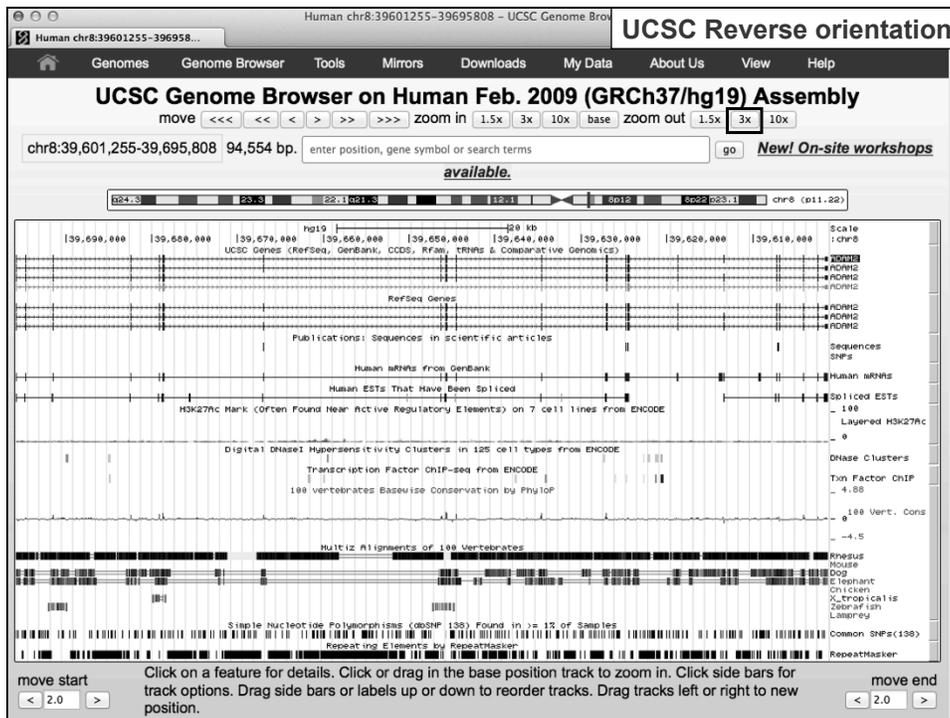
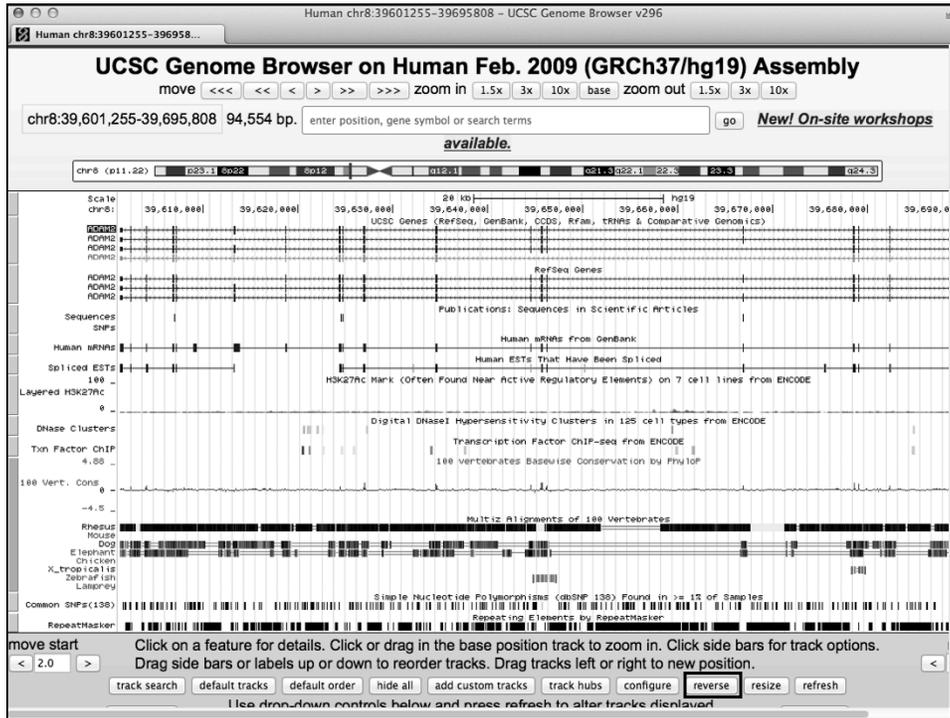
```
>hg19_refGene_NM_001464 range=chr8:39695809-39696808 5'pad=0 3'pad=0
taaatgtacagctcttgaaaactgtcggaagatctcacaacacata
ccctgtgctcgcgaacttcaoctcctgaaatatacaacagtagaactcctt
acttatacaccaaaagccatgtagaagaatctttagctaaatattat
tttaaatagctggaacaataaacaacaataatcattacacagtaaat
ggaacacaagtggtgttatatttaagtaattgtaataacacaaatga
ggataaaacagaactatgcttggatgaaccttacaatcattcttaa
aagaaccagacatgaaagagtagatggtgatgctctacttgcgaaaa
gtcaaaaacagacaaaagaatcttggctttagaagtcattggtgtg
gaggttggaactcggggattgggtgctcttttcaattctctcctgtg
gtactagtacgctgttttttttccactgaatattaaatgacctgtg
aacttatgattatataacttttttctgtttttgtttctgtcttttt
ttttttttttttttttttttgacggaatttcgctctctcaccaggctg
gagtgcaagtgtaagctctttctgctcactgcaacctctgctcctaggt
tcaagcgattctcctgctcagctctccagtagctgggattcaggcac
ccgccaccatgctgctaatttttttgtatttttagtacagagggg
ttcaccatgttgogaaactggtctcgaactcctgatacgtgtttat
atatcaattgaaatttacttaagaaggtttataaattctcctgttcc
tcagctgttgaaagtatttgggtgtgctgtgcttaattaggatca
cgctcagtgagtgctgtctgcaagagacagggtcaggagctcagc
gttccacagcgcaccacccaacctcagcccacctgggctctccagcc
```

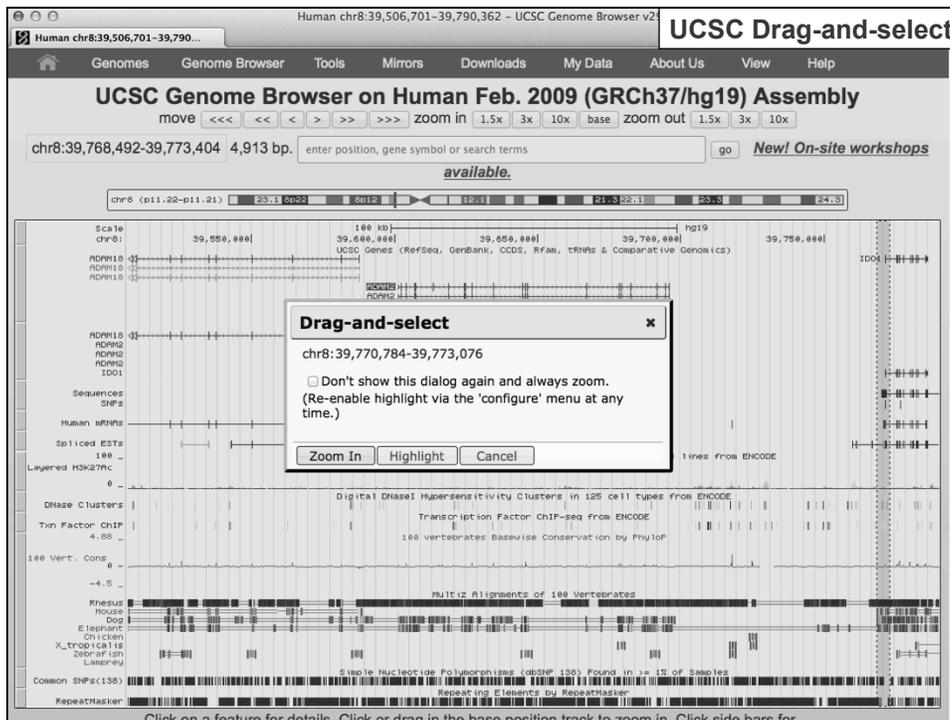
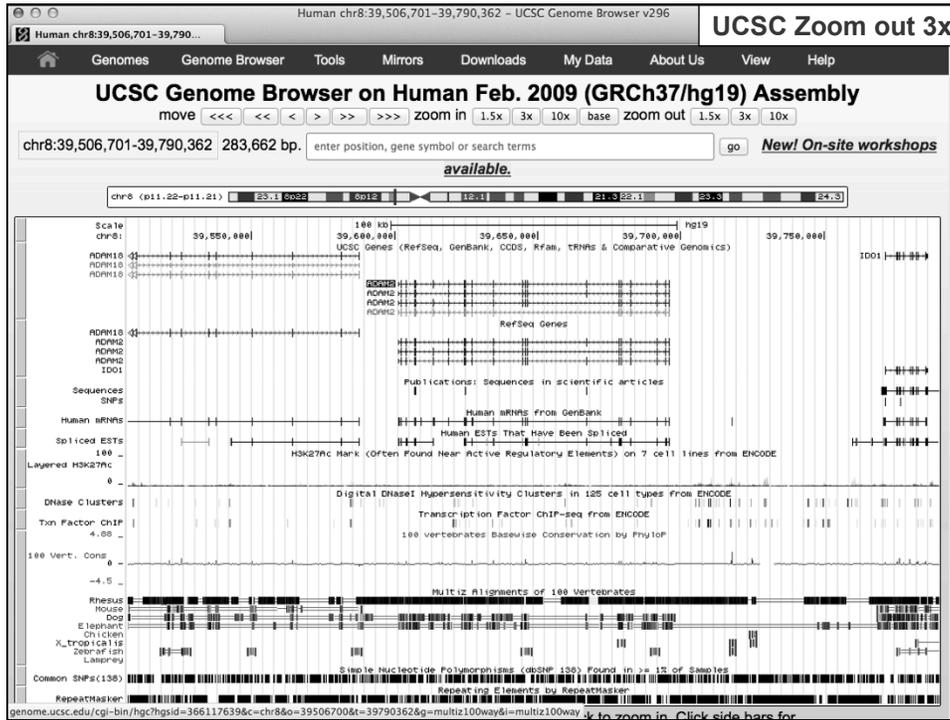
**UCSC**

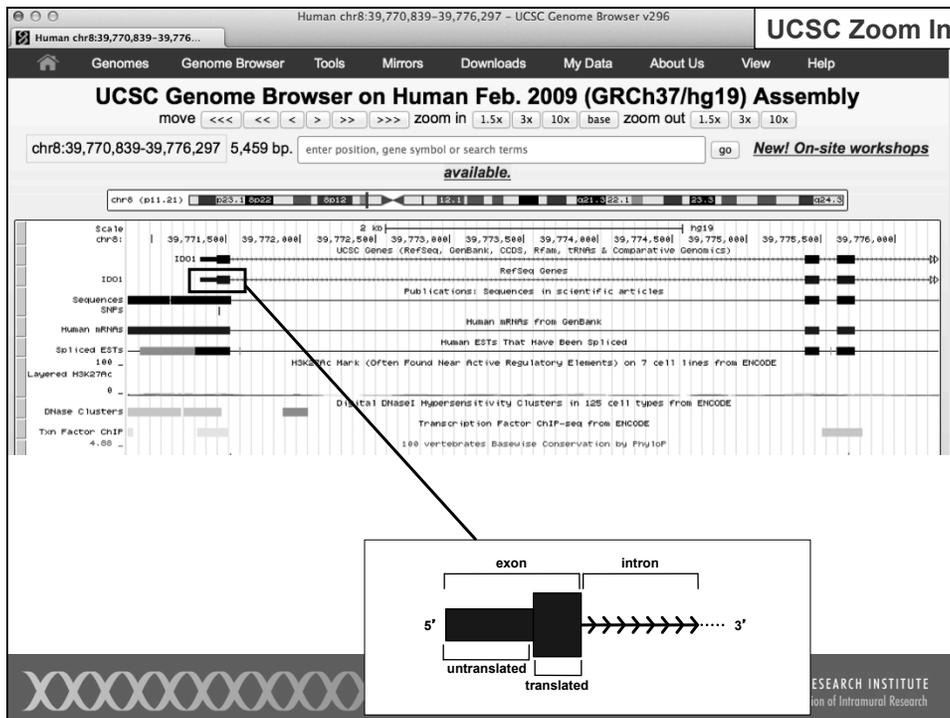
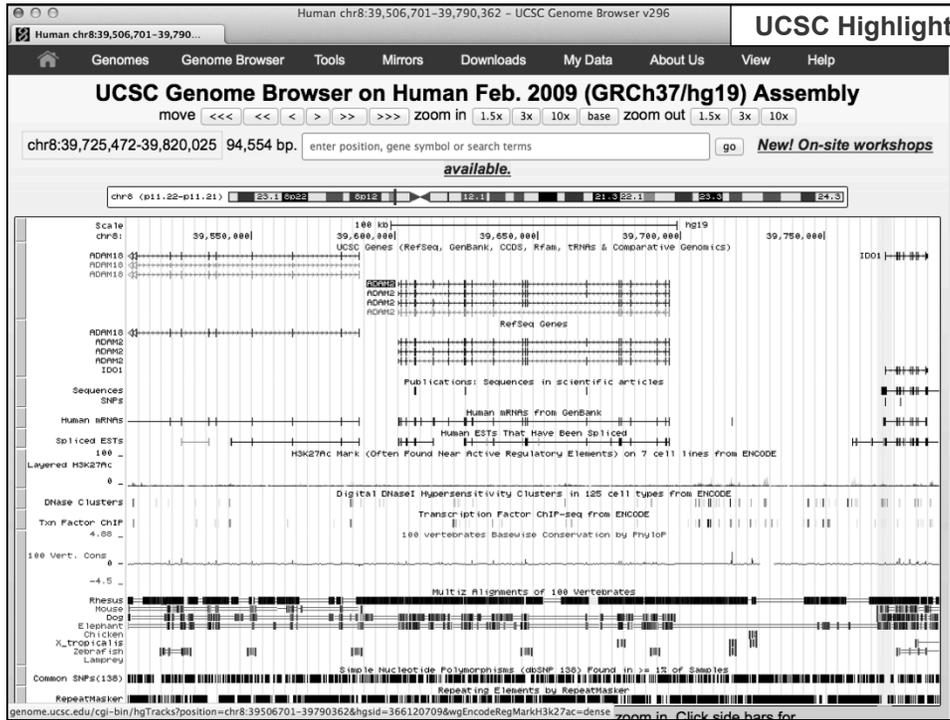
**Navigating around the Genome Browser**

<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	hide	hide	hide
Stanf Nucleosome	SUNY SwitchGear	SwitchGear TSS	TFBS Conserved	TS miRNA sites	UCSF Brain Methyl
hide	hide	hide	hide	hide	hide
UMMS Brain Hist	UW Repli-seq	Vista Enhancers			
hide	hide	hide			
<b>Comparative Genomics</b> refresh					
Conservation	Cons 46-Way	Cons Indels MmCf	Evo Cpg	GERP	phastBias gBGC
full	hide	hide	hide	hide	hide
Primate Chain/Net	Placental Chain/Net	Vertebrate Chain/Net			
hide	hide	hide			
<b>Neandertal Assembly and Analysis</b> refresh					
<b>Denisova Assembly and Analysis</b> refresh					
<b>Variation</b> refresh					
Common SNPs(138) dense	1000G Ph1 Accsbl	1000G Ph1 Vars	All SNPs(135)	All SNPs(137)	All SNPs(138)
hide	hide	hide	hide	hide	hide
Common SNPs(135)	Common SNPs(137)	DGV Struct Var	Flagged SNPs(135)	Flagged SNPs(137)	Flagged SNPs(138)
hide	hide	hide	hide	hide	hide
Genome Variants	GIS DNA PET	HAIB Genotype	HapMap SNPs	HGDP Allele Freq	Mult. SNPs(135)
hide	hide	hide	hide	hide	hide
Mult. SNPs(137)	Mult. SNPs(138)	NumtS Sequence	Segmental Dups	Self Chain	Mult. SNP/CNV Arrays
hide	hide	hide	hide	hide	hide
<b>Repeats</b> 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  $\geq 1\%$  of Samples** (\*All Variation tracks)

Display mode:  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:  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  Locus  Coding - Synonymous  Coding - Non-Synonymous   
 Untranslated  Intron  Splice Site

[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

ADPKC2  
 ADPKC2  
 ADPKC2

Hedenstierna2008  
 Fuluck1997  
 Chen1999

Sequences in Articles: PubMedCentral and Elsevier  
 Kieble

SNPs in Publications  
 Human mRNAs from GenBank

Human ESTs That Have Been Spliced

Spliced ESTs  
 100

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

Transcription Factor ChIP-seq from ENCODE  
 100 vertebrates: Basewise Conservation by PhyloP

100 Vert. cons.  
 -4.5

Rhesus  
 Mouse  
 Dog  
 Elephant  
 Chicken  
 X\_Tropica116  
 Zebrafish  
 Lampbrush

Simple Nucleotide Polymorphisms (dbSNP 138) Found in  $\geq 1\%$  of Samples

rs34966682 rs573382 rs433956 rs282012 rs282095 rs12681344 rs79122716 rs149922990 rs6474177 rs1449799 rs726428245 rs72673685  
 rs122098 rs433959 rs282011 rs18961156 rs7267999 rs2120991 rs39743271 rs7468279 rs731487961 rs275695281 rs2113821  
 rs122989 rs673359 rs7817718 rs3353759 rs282003 rs1947319 rs1451741 rs898472 rs18498917 rs359535824 rs11764464 rs73673829  
 rs11126783 rs420351 rs7817903 rs74784001 rs282002 rs19189159 rs5474176 rs7594 rs7453443 rs24979528 rs726428245  
 rs143373254 rs417869 rs7814938 rs10896756 rs2293081 rs76659751 rs1901386 rs28 rs726428245 rs6851121  
 rs15934991 rs372999 rs453419 rs76776344 rs1901387 rs13265465 rs2122999 rs73 rs7453516 rs6851121  
 rs1805234 rs28664566 rs1572163 rs282010 rs74652324 rs1549565 rs5991336 rs2 rs726428245 rs6851121  
 rs281977 rs9995209 rs12546336 rs1222992 rs13255131 rs3779719 rs6993989 rs73 rs726428245 rs6851121  
 rs281976 rs79433944 rs18089823 rs282009 rs282011 rs6987114 rs145494881 rs73 rs726428245 rs6851121  
 rs14795549 rs12541891 rs59233101 rs7898929 rs7359566 rs7822312 rs7387353 rs111418955 rs117115999 rs13285763  
 rs142876623 rs1104037 rs110118518 rs282008 rs117601926 rs7986936 rs7306549 rs73 rs111418955  
 rs281975 rs77184951 rs18182783 rs282007 rs1886962 rs1886962 rs14982 rs15124453 rs926 rs13238159  
 rs7813282 rs4733922 rs11678387 rs282006 rs74077196 rs1451745 rs1166879 rs11287 rs3419387  
 rs281974 rs330539 rs2844871 rs5991669 rs15943499 rs113541240 rs1101114 rs1899148 rs28653468  
 rs281973 rs25935433 rs282015 rs75674773 rs13252673 rs59821148 rs3912324 rs1739986 rs116226147 rs117115999 rs13285763  
 rs281972 rs2398662 rs282014 rs282004 rs14552946 rs1451744 rs58431931 rs79582812 rs168031873 rs11449616  
 rs281971 rs2398661 rs7836159 rs7464972 rs4436527 rs116775014 rs116775014 rs14543607 rs14536254 rs115764908 rs13345236  
 rs142717481 rs12548521 rs282013 rs1986578 rs18186478 rs28032199 rs73615919 rs141420786 rs10188537 rs147657615  
 rs11599783 rs28200731 rs7379883 rs28982091 rs7379791 rs13376259 rs28118634 rs997296 rs1999148 rs28653468  
 rs116833 rs18993751 rs11777615 rs1895856 rs51297878 rs1811986 rs149927514 rs78432885 rs62511289  
 rs281978 rs188797616 rs111044793 rs2358463 rs11282246 rs112725783 rs141182487 rs141529783 rs101663796  
 rs281978 rs1887294 rs111257991 rs78222544 rs289864576 rs17896693 rs281288034 rs12677911 rs117484296  
 rs7793623 rs114747845 rs112473129 rs113337489 rs78718899 rs148683951 rs14839818 rs11929545 rs114647659  
 rs281969 rs1546336 rs12548241 rs117811917 rs143561399 rs7798895 rs115203943 rs145323361 rs11663110  
 rs5 rs44467 rs14758595 rs88836244 rs14419724 rs149322982 rs145734921 rs145734921  
 rs794 rs4058 rs72642824 rs12785197 rs12785197 rs115857781 rs115857781 rs149323361 rs28681861  
 rs1476 rs63614 rs76387989 rs146791755 rs18115392 rs11015392 rs13285377 rs13285377  
 rs1468 rs99826 rs28564328 rs7819929 rs12674643 rs12674643 rs12674643 rs7947886  
 rs146 rs99826 rs28564328 rs7819929 rs12674643 rs12674643 rs12674643 rs1193525 rs1193525  
 rs28 rs1887426 rs78557624 rs111718283 rs112717748 rs112717748 rs112717748 rs6142932 rs6142932  
 rs142  
 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142 rs142

synonymous SNP  
 non-synonymous SNP

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

**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  
dense hide hide hide hide

**Integrated Regulation from ENCODE Tracks** (\* All Regulation tracks)

Display mode: show Submit

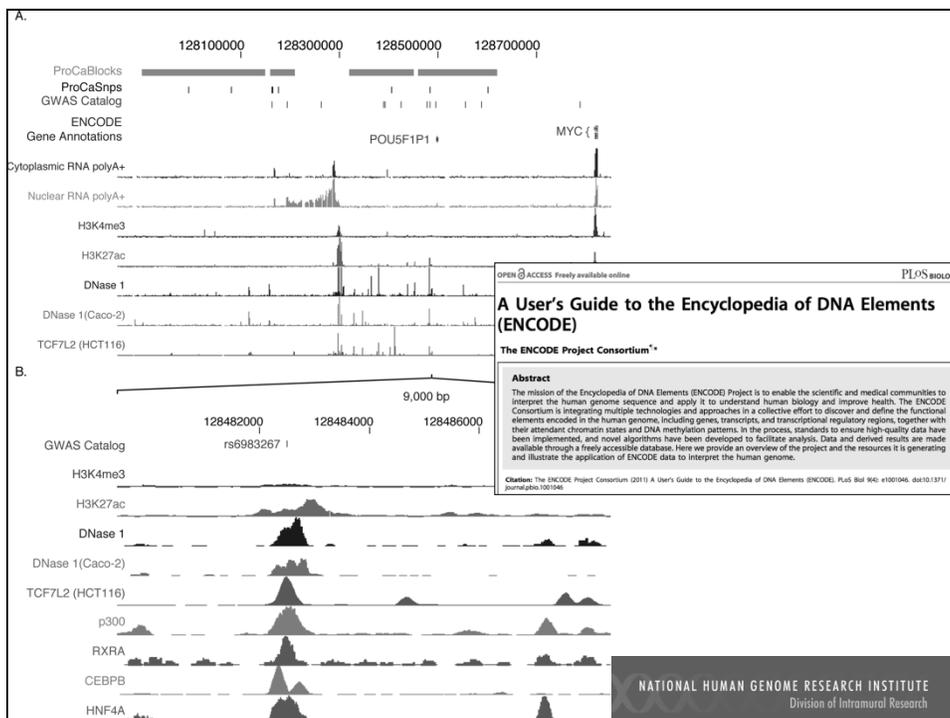
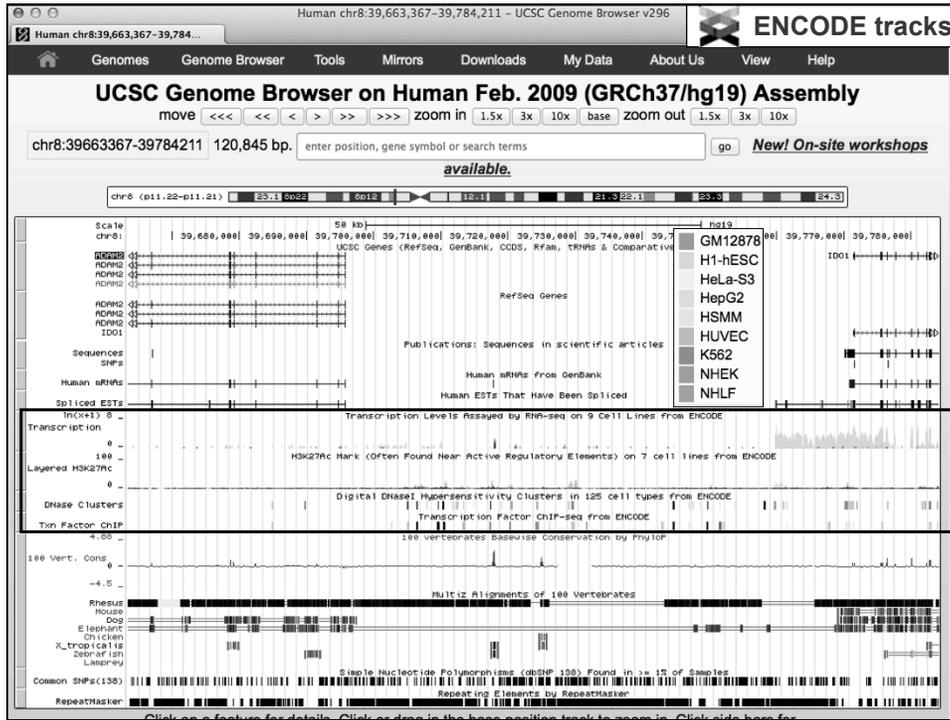
All

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

hide hide hide hide hide hide

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 Methy...
- UMMS Brain Hist
- UW Repli-seq
- Vista Enhancers



**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

disintegrin and metalloproteina... NCBI Resources How To Sign in to NCBI

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]  
MWRVFLLSGLGRLMDSNFDLSPVQITVPEKIRSIKEGIESQASYKIVIEGKPYTVNLQKRNFLPHNF  
RVYSYSGTGMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLFQFENVSYGIEPLESSVGFELHI  
YQVRHKAADVSLNERDIESRDLSPKLSVPEQQDFAKYIEMHVI VEKQLYNMSDSTTVVAKQVFLIG  
LTMALFVDFMTTILSSLELWIDENKIATGAEANELLHFLRWKTSYLVLPAPDAPFLLYREKSNVQDA  
TFQKMKDANYAGGVNLHPRITISLESIAVLAQLLSMGIYDDINKQCQSGAVICINHPFAIHFSGVKI  
FSNCSFEDFAHFISKQKSQLNQPRLDFFKQAVQCNKLEAGECCDQTEQDCALIGETCCDIATCR  
FKAGSNCAEGPCENCLFMSKERMRPSEEDCLPEYCNGSSASCENHYVOTGHPGCLNQMICIDGVCN  
SGDKQCTDTFGKEVEFGPSECYHLNSKTDVSGNCGISDSGYQCEADNLQCKLICKYVGFLLQIPRA  
TIIYANISGHLIAVEFASDHADSQKMKIKDQTSKSGSNKRCNRCVSSSYLGYDCTTDKCNDRGVCNNK  
KHCHSASVLPDSCVQSDLWFGSIDSNGFPVVAIPARLPERRYIENIYHSKMRWPFLEIPFFLIIFC  
VLIAMVKNVFRKKRWRTEDYSSDEQPESESEPKG

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 protein [Homo sapiens]
MWRVFLLSGLGLRMDNFDSLEVOITVPEKTRSLIKEGIESQASYKIVIEGKPYTVNLMQKNFLPHNF
RVYSVSGTGMKPFIDDFQNFCHYQYIEGPKSVVMVSTCTGLRGVLOFENVSVGIEPLESSVGFPHVI
YQHKKADVSLYNEKDIERSDLSPKLSQVPEPQDFAKYIEMHVIIEKOLYNHMGSDTVVAQKVFOLIG
LITNAIFVSNITIISSLELWIDENKLIATGEANELHTFLRWKTSYLVLRFHDVAFLLVYREKSNYVGA
TFQRCMDANTAGOVVLRPRTISLSEAVILAQLLSMGIYDDINKQCSGAVCIHMPFAIFESGVKI
FENCSREDFAHFISKQKSCILNPNRDPFFKQAVCNAKLENGECCCKCFEPCALIGECDCIACRC
FKAGSNCAEGPCENCLFMSKERMCRPSEECDFEYCGNSSASCENHYVOTGHPCGLNWCIDGVCM
SGDKQCTDPTFGKEVEFGPSECYSHLNSKTQVSGNCGISDSGYTQCADNLQCKLICKYVKGFLLOIPRA
LQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
TIIYANISGLICIAVEFASDHADSQKMMIKDGTSCGSKVCRNORCVSSSYLGYDCTDKCNRDGVCKNK
RHCRCASLYLPDCSVQSDLMFQGSIDSGNEFFVAIPARLFERRITENIYHSKFRWFFLEIFPFIIFC
VLAIAIWRVWPKKWRTRDVSSEKWRSESEPRG
```

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	
<a href="#">browser</a>	<a href="#">details</a>	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.**

chr22 22

Scale chr22: 2,453,105 | 59 bases | 2,453,290 | 2410a14 | 2,453,290

Gap

Your Sequence from Blat Search

Non-Chicken RefSeq Genes

RefSeq Genes

Chicken RefSeqs from GenBank

Chicken ESTs That Have Been Spliced

Spliced ESTs

House (Dec. 2011 (GRCm38/mm18)) Chained Alignments

House (Dec. 2011 (GRCm38/mm18)) Alignment Net

Common SNPs (136)

Simple Nucleotide Polymorphisms (dbSNP 136) Found in >= 1% of Samples

RepeatMasker

Repeating Elements by RepeatMasker

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	
<a href="#">browser</a>	<a href="#">details</a>	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  
 Chicken.chr22  
 block1  
 block2  
 block3  
 together

**NP\_001455.3**

```

                mrvlfillg lgglrmdanf dsipvqitvp ekiraiikeg iesqasykiv iegkpytvnl 60
                mqknflphnf rvyssygtgi mkpldqdfgn fchygyyieg ypksvmvat ctglrgvlqf 120
                envsygiepl essvgfehvi yqkhhkadvslynekies rdisfklqsv epqgdakyl 180
                emhvivekql ynhmgsdttv vaqkvfqlig ltnalfvsfn itiillslel widenkiatt 240
                geanelhthf lrwksylvl rphdvafliv yreksnyvga tfqgkmdan yagvvlhpr 300
                tisleslavi laqllslsmg ityddinkcq csgawcimp ealhfsyvkf fencsfedfa 360
                hfiskqkqeg lhnqprldpf fkgavcvgna klsageedc gteqcalig etocdiater 420
                fkagsncaeg pccenclfms kerzcrpsfe ecdlpeycng ssascpenhy vqtghpcgin 480
                qwicidgvcn sgdkqctdf gkevefjpsf cyshlnsktd vsgncglids gytqceadNL 540
                qCGKLiCKv gkfillqpra TIYAnisgH LClavefasd hadsqmwiR DGtsCGenKV 600
                crnqrcvsss ylygdcttdk cndrgvcnkh khochasayl ppdcsvgsdl wpggsidsgn 660
                fppvalparl perryieniy hskpmrwpff lfipffliifc vliainvkvn fqrKkwrted 720
                yssdeqese sepkq
                
```

---

**Chicken.chr22 :**

```

                AATCTGggcT GTGGAAACT CATCTGCaca TAccaaac gqgttcctt caccaaatta 2453164
                aaggtLCCA TCATCTATGC Tcaagtcaa gaaCATCTGT Ggtgtcttt tgatgtaatg 2453224
                catgcacct ccgggacaga tcctctcctg gttAGGATG GCACGaaATG CGTcccgga 2453284
                AAGGTA
                
```

---

**Side by Side Alignment\***

```

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

# UCSC

## Add your own custom tracks

<http://genome.ucsc.edu>



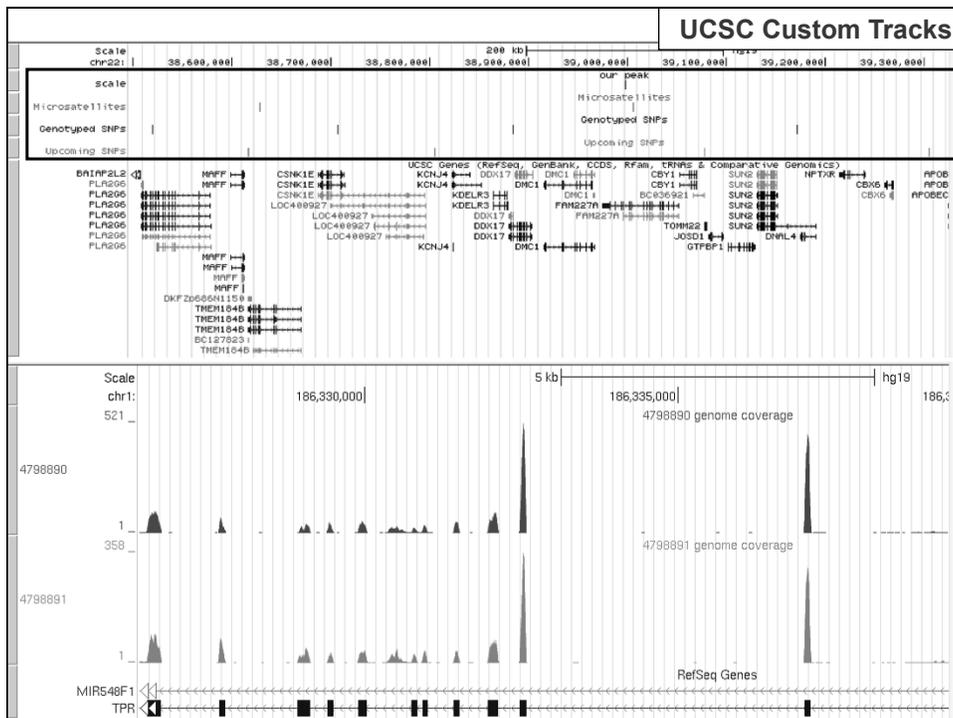
NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
 Division of Intramural Research

## 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](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>



# UCSC

## Table Browser

<http://genome.ucsc.edu>



## 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



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:  
 intersection:  
 correlation:  
 output format: sequence Send output to Galaxy GREAT  
 output file:  
 file type returned: plain text  
 get output summary/stat

**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 r

**Sequence Formatting Options:**

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

get sequence cancel

**UCSC Table Browser:  
200 nt upstream of  
each RefSeq gene**

```

>hg19_refGene_NM_032291 range=chr1:66999625-66999824 5'pad=0 3'pad=0 strand++ repeatMasking=none
ggaaaggcgtgtgtatcttggtaacaaggcgggagcgtgggtggagg
gaatggggacgggaataggtctgtgtctctccggggatctgtgca
ggagatgcaggctggctaccatgtgacgggtccaagctgaaggattg
ccggaggaaggcgggggtgacagctcggccagctgcccctctctccc
>hg19_refGene_NM_032785 range=chr11:50489627-50489826 5'pad=0 3'pad=0 strand-- repeatMasking=none
gtgcgcccaagaaagcgcattcctacagacctgcccctgtgctctag
gggtggcggagcctgctcagttgocaggatgcggtggatcggttgcggag
gagcagocccggggoccaaagtcocagctcagcccggtctgcgcgcg
ggggcggggccttggagcagctcctccagcggctccctccggcctc
>hg19_refGene_NM_001145278 range=chr1:16766967-16767166 5'pad=0 3'pad=0 strand++ repeatMasking=none
agaccacgggttacagagggtctgtccatggcggggcagggcgcttc
tttgcgtcggagggtgtctggaggaaaggagaacctctggaggaggga
gaagcctccggagggtgcocgcacgtgtcttgagocgggtttccagcag
agggcgcacaacgaggcgtgtgaggcccgagctgcccctagccc
                
```

The banner features the Ensembl logo at the top, followed by the text "Variant Effect Predictor (VEP)" in a bold, black font. Below this, the URL "http://www.ensembl.org" is displayed in a smaller, italicized font. At the bottom of the banner, there is a decorative DNA double helix pattern on the left and the text "NATIONAL HUMAN GENOME RESEARCH INSTITUTE" and "Division of Intramural Research" on the right.

The screenshot shows the Ensembl Genome Browser homepage. At the top, there is a navigation bar with links for "BLAST/BLAT", "BioMart", "Tools", "Downloads", "Help & Documentation", "Blog", and "Mirrors". A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a main search area with a "Search:" label, a dropdown menu for "All species", a text input field, and a "Go" button. An arrow labeled "click" points to the "Go" button. Below the search area, there are several sections: "Browse a Genome" with a description and a "Popular genomes" list including Human (GRCh37), Mouse (GRCm38), and Zebrafish (Zv9); "ENCODE data in Ensembl"; "Variant Effect Predictor" with a "VeP" logo; "Gene expression in different tissues"; "Find SNPs and other variants for my gene" with a sequence example; "Retrieve gene sequence" with a sequence example; "Compare genes across species"; "Use my own data in Ensembl"; "Learn about a disease or phenotype"; "What's New in Release 75 (February 2014)" with a list of updates; "Latest blog posts" with a list of recent posts; and "Did you know...?" with a link to a workshop. At the bottom, there is a footer with the Sanger logo and text about the project's funding and goals.

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

Ensembl: VEP tyra@nhgri.nih.gov

Human (GRCh37) Jobs

Tools  
 Web Tools  
 Variant Effect Predictor  
 Configure this page  
 Manage your data  
 Export data  
 Bookmark this page  
 Share this page

### Variant Effect Predictor

**New VEP job:**

**Input**

Species:

Assembly: GRCh37

Name for this data (optional):

Input file format (details):

Either paste data:

```
rs35935433
rs144646998
rs145143599
rs34417912
```

Or upload file:  No file selected.

Or provide file URL:

Or select previously uploaded 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

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

Ensembl: VEP tyra@nhgri.nih.gov

Human (GRCh37) Jobs

Tools  
 Web Tools  
 Variant Effect Predictor  
 Configure this page  
 Manage your data  
 Export data  
 Bookmark this page  
 Share this page  
 Download view as CSV

### Variant Effect Predictor results

Summary statistics for ticket 1NXgvCNO5OblgIob:

Category	Count
Variants processed	4
Variants remaining after filtering	4
Novel / existing variants	0 (0.0%) / 4 (100.0%)
Overlapped genes	2
Overlapped transcripts	5
Overlapped regulatory features	-

**Consequences (all)**

● synonymous\_variant: 33%  
 ● missense\_variant: 33%  
 ● intron\_variant: 22%  
 ● downstream\_gene\_variant: 11%

**Coding consequences**

● synonymous\_variant: 50%  
 ● missense\_variant: 50%

**Results preview**

Navigation

Showing 18 results for variants 1-4 of 4 | Show 1 All |

Uploaded variation : is : defined

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	AC136365.1	-	-	-	miRNA	-	-	-	-
-	-	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:

HGVs names: This variation has 7 HGVs 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**

5.00 kb

Contigs  
 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**

Variation legend	<ul style="list-style-type: none"> <li>Splice acceptor variant</li> <li>Missense variant</li> <li>Mature miRNA variant</li> </ul>	<ul style="list-style-type: none"> <li>Stop gained</li> <li>Splice region variant</li> <li>Non coding exon variant</li> </ul>	<ul style="list-style-type: none"> <li>Frameshift variant</li> <li>Synonymous variant</li> <li>Intron variant</li> </ul>
------------------	---	---	--

Share this page

Download view as RTF

BLAST this sequence

**Key**

Variations

Intronic	Mature miRNA	Missense
Non-coding exon	Splice acceptor	Splice region
Synonymous		

• **Focus variant**

```

TCASCCTGGCCAACTGTGAAACTCCGCTCTACTAAAAATACAAATTAGCGGGTG
TGGTGGTACCACCTGTAAATCCAGCTACTCAAGAGGCTGAGGACAGGAAATCACTTGAAC
CTGGTGGCAGAGCTGCACTGAGCGAGATCACACACTGCACTTACCCAGGGCANGA
GTGACACTCACTCAAAAAGGGCCAGTCTAAAATATTAGGAAATTAATGGATATAA
AGACTACCTACCTTATTTGAACCACAGAGTCCACCTTTATCCAAATCTTTGGCTG
TCTGCATGATCACGGCAATTCCACAGCAATGAGAGATGTCCTTATGTTGGCATAA
ATAATAGTGCCTCTGGAAATTTGAAATAAAATTTACCTATATATTACAATTAATTTT
CCCACTGCAGATTGATAATTTATCCACAATAAAAATTAATTTAATTTAGAGA
AGTACACAGATATAATTTTAAAATAGATGATTAATAAGTATTCAAGGTTCCAG
CTAATACATGCAATTTTAAAACACCAAAAAGAGCCCTGCCAAATTCCTATAA
TAACTGAACATTTTAGTAATTAATTAAGCACTAAGTTGGTGCAAAATAATAGTGGTT
TTGCCACTTTTAAATGGCAAAAGCCCAATCACTTTGCACCACTAATATCCCTAG
TTASCCTGCACACTTTTTTTGTGATTCCTTCACTTTTGAATCTGAAACAAAATCAA
ACCCAGCTCAATGAATTA
                    
```

**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

**Ensembl Location tab: Region in detail**

Region in detail

Chromosome bands

Gene Legend

Location: 8:39613246-39613446 Gene:

Chromosome bands

Sequence

Genes (GENCODE...)

CCDS set

Human cDNAs (R...)

Variation Legend

Information

Ensembl genome browser 75: Homo sapiens - Region in detail - Chromosome 8: 39

**Ensembl Location tab: Configure page**

Region in detail

Configure Region Image

Configure Overview Image

Configure Chromosome Image

Personal Data

Genes and transcripts (11/85)

mRNA and protein alignments (1/14)

Human BodyMap 2.0

Adipose

Adrenal

Blood

Brain

Breast

Colon

Heart

Kidney

Liver

Lung

Lymph

Merged (incl.Pooled)

Ovary

Pooled

Prostate

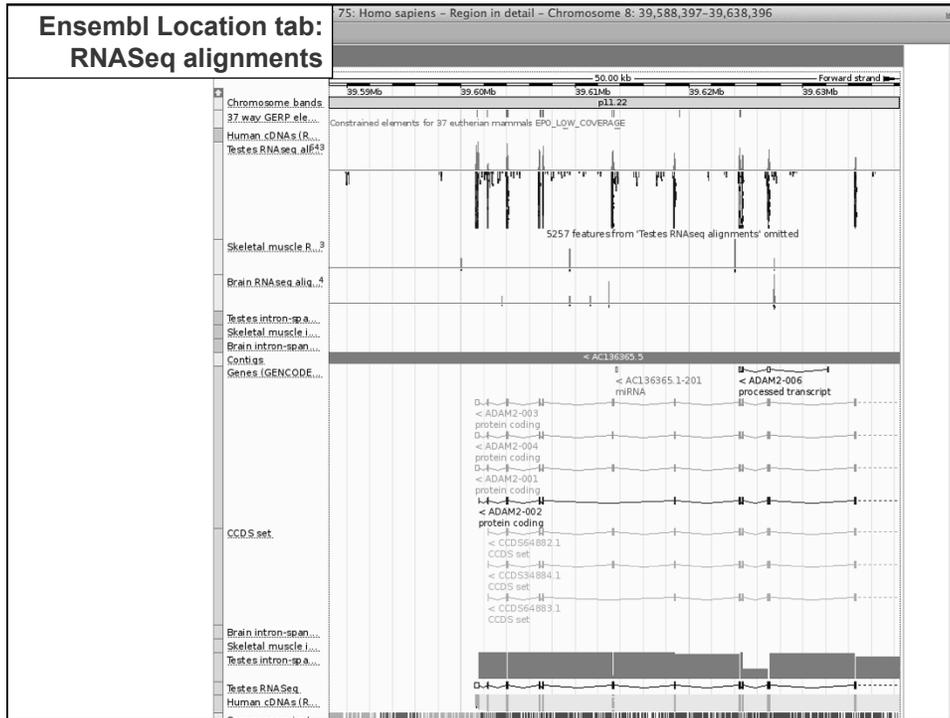
Skeletal muscle

Testes

BAM files

Gene models

Intron-spanning reads



**Ensembl**

**Gene tab**

<http://www.ensembl.org>



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

### Ensembl Location tab: Region in detail

Location: 8:39613346-39613446  
 Gene:

Chromosome bands: p11.22

Genes (GENCODE 19):

- < ADAM2-003 protein coding
- < ADAM2-001 protein coding
- < ADAM2-000 protein coding
- < ADAM2-002 protein coding
- < ADAM2-004 protein coding
- < ADAM2-005 protein coding
- < ADAM2-006 protein coding
- < ADAM2-007 protein coding
- < ADAM2-008 protein coding
- < ADAM2-009 protein coding
- < ADAM2-010 protein coding
- < ADAM2-011 protein coding
- < ADAM2-012 protein coding
- < ADAM2-013 protein coding
- < ADAM2-014 protein coding
- < ADAM2-015 protein coding
- < ADAM2-016 protein coding
- < ADAM2-017 protein coding
- < ADAM2-018 protein coding
- < ADAM2-019 protein coding

CCDS.set:

- < CCDS: Gene type Known protein coding
- < CCDS: Transcript type Known protein coding
- < CCDS: Strand Reverse
- < CCDS: Base pairs 2,672
- < CCDS: Amino acids 735

Human cDNAs (RefSeq):

- < RefSeq: Transcript
- < RefSeq: Protein
- < RefSeq: CDS

Gene Legend:

- Merged Ensembl/Havana
- Pseudogene
- Processed transcript
- RNA gene

Ensembl genome browser 75: Homo sapiens - Summary - Gene: ADAM2 (ENSG00000104755)

### Ensembl Gene tab: Gene summary

Human (GRCh37) Location: 8:39,613,346-39,613,446 Gene: ADAM2 Transcript: ADAM2-001 Variation: rs3959354

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 by
- Orthologues (54)
- Paralogues (6)
- Protein families (1)
- Phenotype
- Genetic Variation
- Variation table
- Variation image
- 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	CCDS34484	Y
ADAM2-003	ENST00000347580	2535	ENSP00000347580	716	Protein coding	CCDS64882	Y
ADAM2-002	ENST00000521980	2125	ENSP00000429352	672	Protein coding	CCDS64883	Y
ADAM2-004	ENST00000378953	2125	ENSP00000369162	579	Protein coding	-	Y
ADAM2-005	ENST00000523181	728	No protein product	-	Processed transcript	-	Y
ADAM2-006	ENST00000520434	520	No protein product	-	Processed transcript	-	-

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: [CCDS34484](#), [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

Species set	Show details	1:1	1:many
Primates Humans and other primates	<input type="checkbox"/>	8	0
Rodents Rodents, rabbits and related species	<input type="checkbox"/>	7	0
Laurasiatheria Carnivores, ungulates and insectivores	<input type="checkbox"/>	14	0
Placental Mammals All placental mammals	<input type="checkbox"/>	34	0
Sauropsida Birds and Reptiles	<input checked="" type="checkbox"/>	2	5
Fish Ray-finned fishes	<input type="checkbox"/>	0	0
All All species, including invertebrates	<input type="checkbox"/>	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]	<ul style="list-style-type: none"> <li>Region Comparison</li> <li>Alignment (protein)</li> <li>Alignment (cDNA)</li> <li>Gene Tree (image)</li> </ul>	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: R4GDHS]	<ul style="list-style-type: none"> <li>Region Comparison</li> <li>Alignment (protein)</li> <li>Alignment (cDNA)</li> <li>Gene Tree (image)</li> </ul>	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]	<ul style="list-style-type: none"> <li>Region Comparison</li> <li>Alignment (protein)</li> <li>Alignment (cDNA)</li> <li>Gene Tree (image)</li> </ul>	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

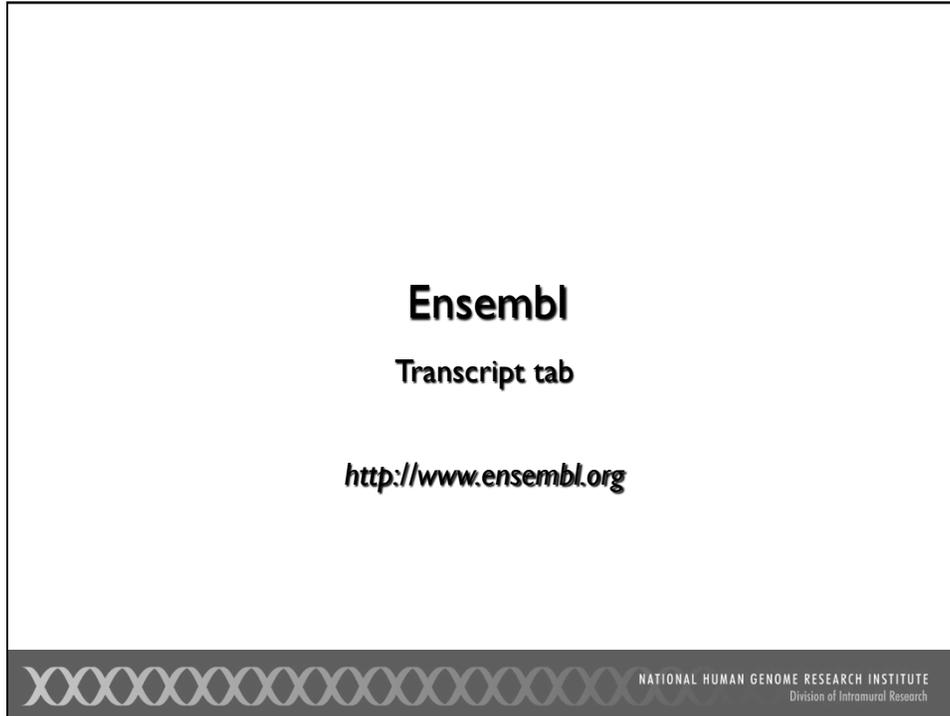
PROSITE profiles

Pfam domain

Prints domain

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



**Ensembl Transcript tab:  
 Transcript summary**

Ensembl genome browser 75: Homo sapiens - Summary - Transcript: ADAM2-001

Human (GRCh37) Location: 8:39,613,254-39,613,446 Gene: ADAM2 Transcript: ADAM2-001 Variation: rs95935433

**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 [ENSG00000104755](#)  
 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  
 This transcript is a member of the Human CCDS set: [CCDS34884](#)

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: [OTT\\_HUMT00000376926](#) (version 1)

Ensembl genome browser 75: Homo sapiens - Supporting evidence - Transcript: ADA...

**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**

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: ADA...

**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

```

MWRVFLLSGLGGLRMSNFDLFPVQITVPEKIRSIIEKESQASRYKIVIEGRPYTVNL
MQKNFLPHNFRVYSYSGTGIMKPLDQDFNCFYQYIEGPKSVVMVSTCIGLRGVLP
ENVSYGIEPLESSVGFHVIYQVKKKADVSLYNEKDIESPDLSPFLQSVFQQDFARYI
EMGVIVKQLYNNMGGSTTVKQKVPQLIGLNAPFVSNFTIISLEIMIDENKIAIT
GEANELLHTFLRKEKTSYLVRPHDDVLLVYREKSNVYVGFQGMCDANVAGGVVLP
TISLESLAVILAQLLSLMSGITYDDINKQCQSGAVIMNPEAIFHSGVKIFNSCSFEFA
HFIKQKQSCQLINQPRLDFFKQAVCNKLEAGEECDCGTEQDCALIGETCCDIATCR
FRAGSNCAEGPCENCLEFMSKERMCRPSFECCDLPYCNSSASCPENHYVQVGHPCCLN
QWICIDVCMGSDGKCTDFTGKEVFEFGFSECHLNSKSDVSGNCSISDSYFCQCEANL
QCKLICKYVVKPLQIPRATIIYANISGHLIAVEFASDHADSQMMKIDGTSKCSNRY
CRNQRCSVSSYLGYDQTDKCDNRGVCNKKHCHCSALFPDQCSQSDLPFGGSDISGN
PPFVAIPARLPERRYIENIYHSKPMRWPFLEFIFEFIIFCVLIAMVKVNFQRKKWRTE
YSSDQPESESEPKG
    
```

Ensembl release 75 - February 2013

[Permanent link](#) - [View in archive site](#)

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

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 page header includes "Ensembl archive" and the user email "tyra@nhgri.nih.gov". The main navigation bar contains "BLAST/BLAT", "BioMart", "Tools", "Downloads", "Help & Documentation", "Blog", and "Mirrors". The left sidebar shows a tree view of genomic data types, with "Transcript" selected. The main content area is titled "View in archive site" and contains a search box and a list of available 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". At the bottom of the page, there is a search box labeled "BLAST this sequence".

**Ensembl**

Find a chicken homolog of a human protein

<http://www.ensembl.org>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research

**Alignment Summary (click arrow to hide)**  
 Select rows to include in table, and type of sort (Use the 'ctrl' key to select multiples)

Query off_name	Subject off_name	Chromosome off_name	Scaffold off_name	Contig off_name	Stats off_name	Sort By
Start	Name	Name	Name	Name	Score	>Contig
Start	Start	Start	Start	Start	E-val	<Score
Start	Start	Start	Start	Start	%ID	>Score
Start	Start	Start	Start	Start	Length	
[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	26726477 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:8				
[A] [S] [G] [C]	444 501 +	Chr:22				
[A] [S] [G] [C]	339 443 +	Chr:22				

Query location : ref|NP\_001455.3| 6 to 662 (+)  
 Database location : 15 6235722 to 6237707 (+)  
 Genomic location : 15 6235722 to 6237707 (+)

Alignment score : 1365  
 E-value : 8.4e-134  
 Alignment length : 685  
 Percentage identity: 35.33

Query: 6 FLLSGLGLRMDSN----FDSLPLVQIVPEKIRSIIEKIGESQAS--YKIVIEGKPYTVN 59  
 FL+GL L D + + +I P+K S K G SQ S Y I I+G YI+  
 Sbjct: 6235722 FLLTGLLLCADCSPPQPTWGYTAYEIVTPPKAGS--KAGRASQGSMSYFIIIGQVNYIHH 6235895

Query: 60 L-MQRNLFPHNFRVYSYSGTGIMKFLDQDFQNFCHYQVIEGYPKSVVMVSTCTGLRGLV 118  
 L +K F+ NF + + G + C+Y+G+EG S V ++TC+GLRG+L  
 Sbjct: 6235896 LRHKRGFVKNFPLTRDSEGVMEQPRVLADCYIHGVEGILDSTVTLTTCGRLGLL 6236075

Query: 119 QFENVSYGIEPLESSVGFHVIQVHKHKADVSLYNEKDIERSDLFKLQSVPEQDP-- 176  
 Q +N+SY IEP L+S FEH++ Q + +Y K++ F + P+Q F  
 Sbjct: 6236076 QIGNLSYSIEPLAASSTFHEHLLQREAVVPGTVIY--KTLQGR-APFGRGTAPRQ-POP 6236243

Query: 177 ---ARYEMHVIKQLYNHMGSDTTVAQKVFQILGLTNAIFVFNITILLSSLELWID 233  
 +Y+E V+V+K++ G+ T V +V +I L + +F S + ++L+ LE+W +  
 Sbjct: 6236244 WGRTRYLEMLVVVDKEGDFTFGTSITNVTLEVIEIINLVGLFSSVRLVLLTVLEIWT 6236423

Query: 234 ENKIATTEANELLHFLRWKTSYLVRP-HDV---AFLVYRE-KSNYVG--ATPQGR 285  
 +N I+ T ++LH+ F RW+ + HDV A L R ++ +V G + F  
 Sbjct: 6236424 KNPISTKNTIQLVHFNWRIRIQHGFAMHMDVGLFASLDFRSRTRALHVGGSNFASA 6236603

Query: 286 MCDANYAGVVLHPTISLESLESLAVILAQLLSLMSGITYDDKNCQCSGV+ -CINPEAIIH 344  
 C+ ++ VV + +E+ AV +A+ L +G+ +DD C+C A CIMNP++  
 Sbjct: 6236604 -CNRQSSAVVSPFAKHTYIET-AVHVAHELGVVLGMEHDD-EHCRGNGASKCINPKSTV 6236774

Query: 345 FSGVKIFSMCSPEFPAHFSKQSLHNQPR-LDPFFQQAQVCGMAKLEAGECCDCTE 403  
 G FSNCS + + F+ + QCL+N P + F Q+ CGN LE EECDCDTE  
 Sbjct: 6236775 SYG---FSNCSKYDFDFTISGGQGLNIPSSIAFVQR--CGNVLEDRRECDCDCTE 6236939

**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 links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, and Blog. The main content area is divided into a left sidebar and a main query area. The sidebar shows three dataset selection options, all currently set to 'None Selected'. The main query area has a dropdown menu for 'Ensembl Genes 75' and a text input field for 'Danio rerio genes (Zv9)'. Below this, there are sections for 'REGION' and 'GENE' filters. The 'GENE' section is expanded, showing a list of Ensembl Gene IDs with a 'Limit to genes ...' checkbox and radio buttons for 'Only' and 'Excluded'. A 'Browse...' button is visible next to the list. The 'Source (gene)' filter is set to '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:  Structures,  Homologs,  Variation,  Sequences

GENE:
 

- 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 transcript
- Unigene ID
- 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	ENSDART00000147844	16	20493224	20528393	rxrbb		
ENSDARG00000002006	ENSDART00000128914	16	20493224	20528393	rxrbb		XM_005157945
ENSDARG00000002507	ENSDART00000139859	16	16045949	16118555	flga10		XM_005157916
ENSDARG00000002507	ENSDART00000139859	16	16045949	16118555	flga10		XM_003200156
ENSDARG00000002507	ENSDART00000011224	16	16045949	16118555	flga10		
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	ENSDART00000121988	16	15611220	15622320	grwd1	NM_001003509	
ENSDARG00000005762	ENSDART00000138611	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	22955445	22973946	hibadhb		
ENSDARG00000007959	ENSDART00000008429	16	22955445	22973946	hibadhb	NM_201160	
ENSDARG00000007959	ENSDART00000132407	16	22955445	22973946	hibadhb		
ENSDARG00000007959	ENSDART00000131452	16	22955445	22973946	hibadhb		
ENSDARG00000009023	ENSDART00000146436	16	22143616	22239485	ankrd28b		XM_684152
ENSDARG00000009023	ENSDART00000027020	16	22143616	22239485	ankrd28b		
ENSDARG00000013371	ENSDART00000007842	16	14545332	14561307	isoc2	NM_001079953	
ENSDARG00000013371	ENSDART00000146997	16	14545332	14561307	isoc2		
ENSDARG00000018787	ENSDART00000015956	16	28521948	28537442	gflna1b		
ENSDARG00000018787	ENSDART00000035279	16	28521948	28537442	gflna1b		
ENSDARG00000019658	ENSDART00000141032	16	12823290	12882007	pou2f2b		



**Galaxy: Step 1: Download transcript data from UCSC**

Galaxy  
 Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools **Load Data**

search tools

Get Data  
 Upload File from your computer  
**UCSC Main table browser**  
 UCSC Archae table browser  
 EBI SRA ENA SRA  
 BioMart Central server  
 GrameneMart Central server  
 modENCODE fly server  
 modENCODE worm server  
 WormBase server  
 EuPathDB server  
 GenomeSpace Import from file browser

Send Data  
 Lift-Over  
 Text Manipulation  
 Convert Formats

**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 [Open-Helix 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 add custom tracks track hubs

table: knownGene describe table schema

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

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

Galaxy Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

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

History

search datasets

knownGene coding sequence lengths  
 11.5 MB

Tags:  
 Add tags

**Galaxy: Step 2: Extract coding exons**

Galaxy  
 Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools **Load Data**

search tools

Get Data  
 Send Data  
 Lift-Over  
 Text Manipulation  
 Convert Formats  
 FASTA manipulation  
 Filter and Sort  
 Join, Subtract and Group  
 Extract Features  
**Gene BED To Exon/Intron/Codon BED**  
 Fetch Sequences  
 Fetch Alignments  
 Get Genomic Scores  
 Operate on Genomic Intervals

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

Extract  
 Coding Exons only

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

this history item must contain a 12 field BED (see below)

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.

Galaxy Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

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	+

History

search datasets

knownGene coding sequence lengths  
 11.5 MB

Tags:  
 Add tags

1: UCSC Main on Human: knownGene (genome)  
 82,960 regions  
 format: bed, database: hg19

Tags:  
 Add tags

display in IGB Local Web  
 rtenlaw at EMBL Current

2: Gene BED To Exon/Intron/Codon BED on data 1

### 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?:

**Execute**

**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?:

**Remove Operation 1**

**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

The screenshot displays the Galaxy interface for creating a workflow. On the left, a list of tools is shown with checkboxes to include them in the workflow. The tools are: UCSC Main (disabled), Gene BED To Exon/Intron/Codon BED, Compute, and Group. On the right, the workflow steps are defined: Step 1: Input dataset (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), and Step 4: Group (version 2.0.0). A 'Run workflow' button is visible at the bottom right.

## 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/>


 NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
 Division of Intramural Research

## Current Protocols in Bioinformatics

<p><b>The UCSC Genome Browser</b></p> <p>Donna Karolchik,<sup>1</sup> Angie S. Hinrichs,<sup>1</sup> and W. James Kent<sup>1</sup></p> <p><sup>1</sup>Center for Biomolecular Science and Engineering, University of California Santa Cruz, California</p> <p><b>ABSTRACT</b></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 include gene models, repeats, and other genomic features. The UCSC Genome Bioinformatics Group and external collaborators have also provided tracks for microRNA, mRNA and expressed sequence tag alignments, simple nucleotide conservation, expression and regulatory data, phenotype and variation data, and phylogenetic comparative genomics data. All information relevant to a region of the genome is available in a single window, facilitating biological analysis and interpretation. The underlying genome browser tracks can be viewed, downloaded, or analyzed using another Web-based application, the UCSC Table Browser. Users</p>	<p style="text-align: right;">UNIT 1.4</p> <p><b>Using Galaxy to Perform Large-Scale Interactive Data Analyses</b></p> <p>Jennifer Hillman-Jackson,<sup>1</sup> Dave Clements,<sup>2</sup> Daniel Blankenberg,<sup>1</sup> James Taylor,<sup>2</sup> Anton Nekrutenko,<sup>1</sup> and Galaxy Team<sup>1,2</sup></p> <p><sup>1</sup>Penn State University, University Park, Pennsylvania  <sup>2</sup>Emory University, Atlanta, Georgia</p> <p><b>ABSTRACT</b></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? Galaxy is a Web-based platform for performing genomic analyses. Galaxy was designed to be easy to use and is riddled with complexities outside of the core analysis. The authors believe that Galaxy provides a powerful environment for performing genomic analysis in an intuitive Web application, integrating bioinformatics tools previously only available to command-line based environments. We will demonstrate through examples how Galaxy specifically brings together (1) data sources, for example, UCSC’s Eukaryote and Genes tracks, (2) command-line tools (wrapped Unix functions, format converters), and (3) 3rd-party analysis tools. <i>Curr. Protoc. Bioinform.</i> 30:1.15.1-1.15.48. © 2010 by John Wiley &amp; Sons, Inc.</p> <p style="text-align: right;">UNIT 10.5</p>
<p><b>Using the Ensembl Genome Server to Browse Genomic Sequence Data</b></p> <p>Xosé M. Fernández-Suárez<sup>1</sup> and Michael K. Schuster<sup>1</sup></p> <p><sup>1</sup>EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom</p> <p><b>ABSTRACT</b></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 (<a href="http://www.ensembl.org/">http://www.ensembl.org/</a>), 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 &amp; Sons, Inc.</p>	<p style="text-align: right;">UNIT 1.15</p> <p><b>Access from NIH at</b>  <a href="http://onlinelibrary.wiley.com/book/10.1002/0471250953">http://onlinelibrary.wiley.com/book/10.1002/0471250953</a></p>

Keywords: computer graphics • databases • genetic • genetic variation • genome • genome sequence • homology • genome • genome sequence