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

JOHNS HOPKINS MEDICINE
CONTINUING MEDICAL EDUCATION

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

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

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 assembles

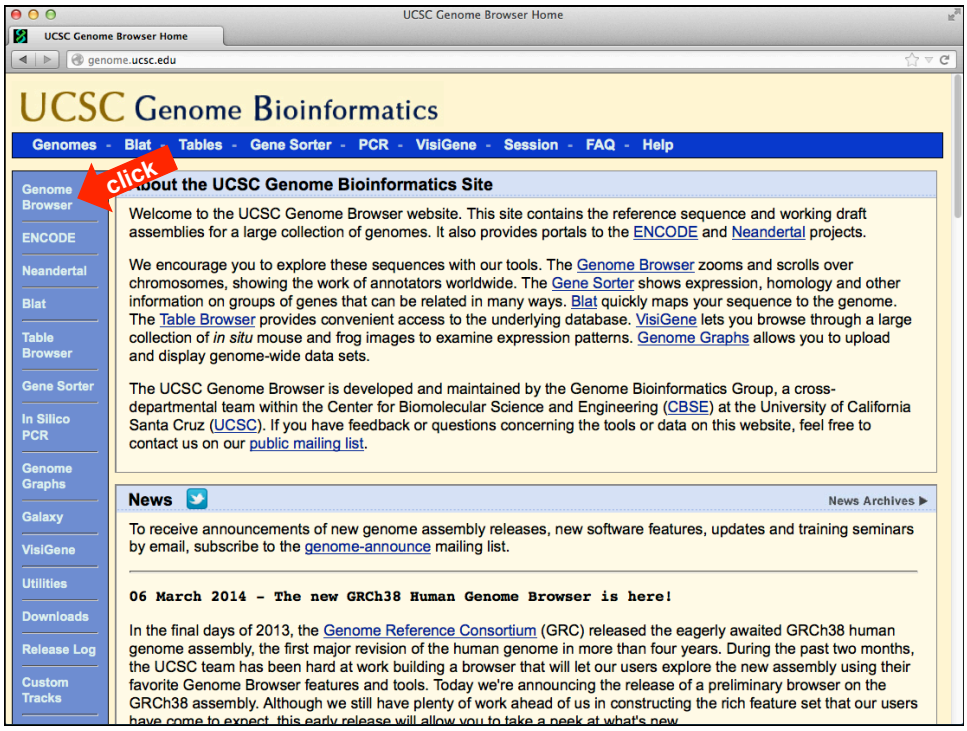
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

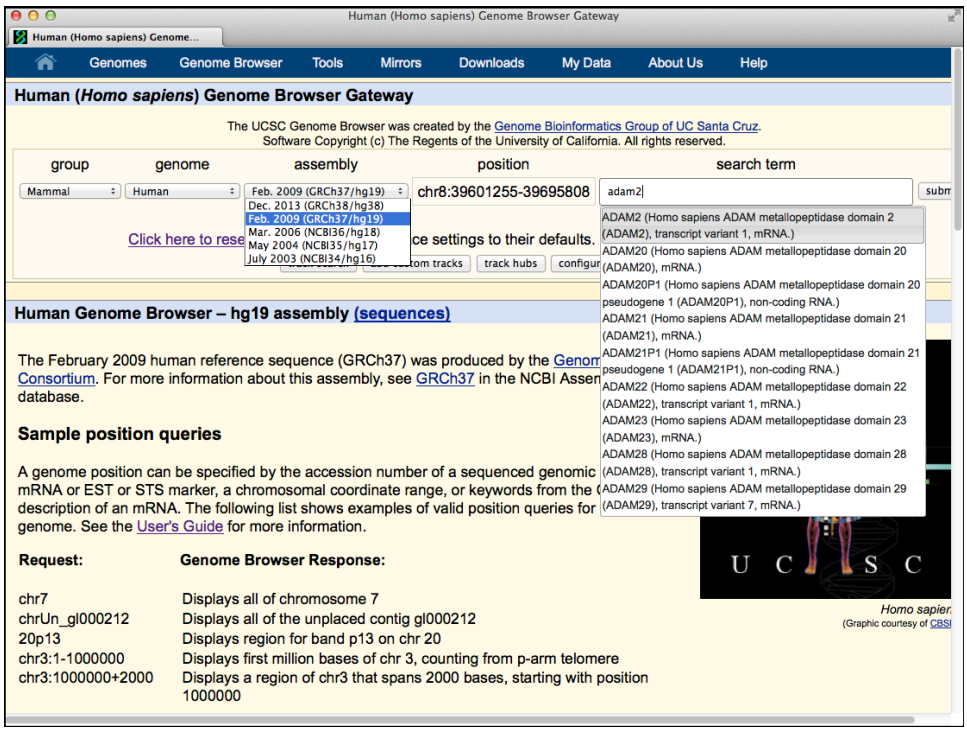
UCSC

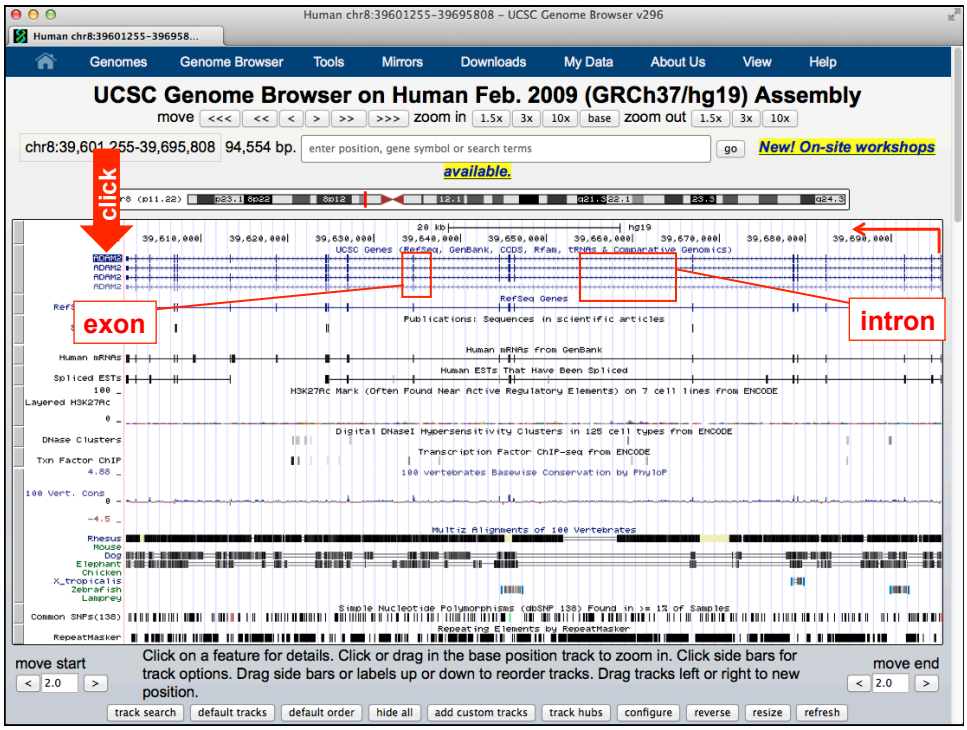
View a region in the genome by querying with a gene symbol

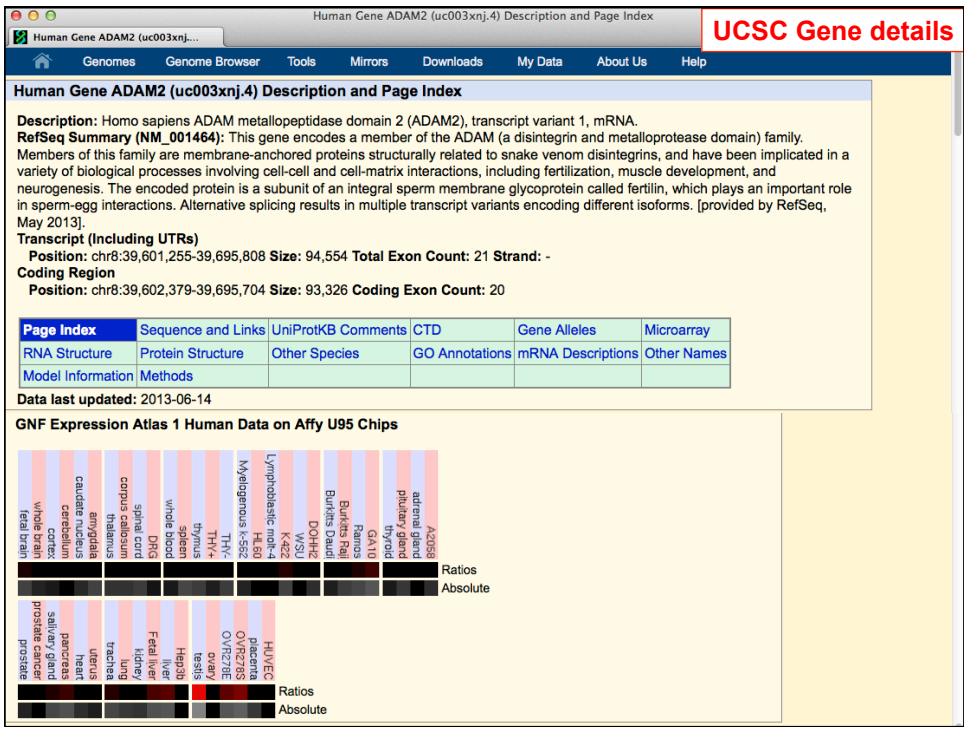
http://genome.ucsc.edu

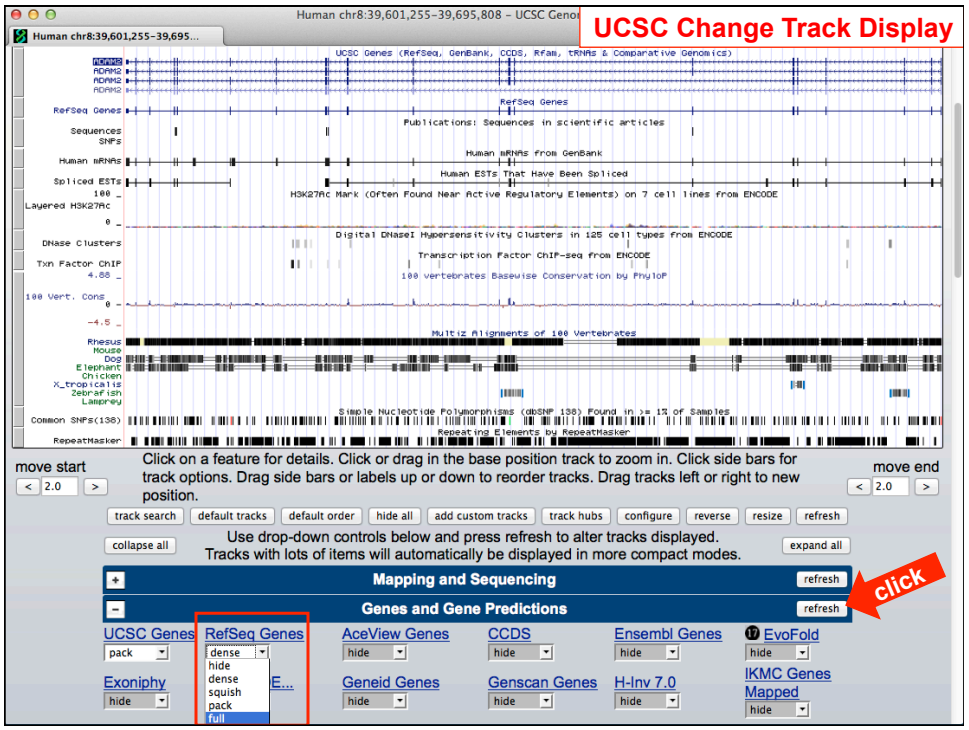
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

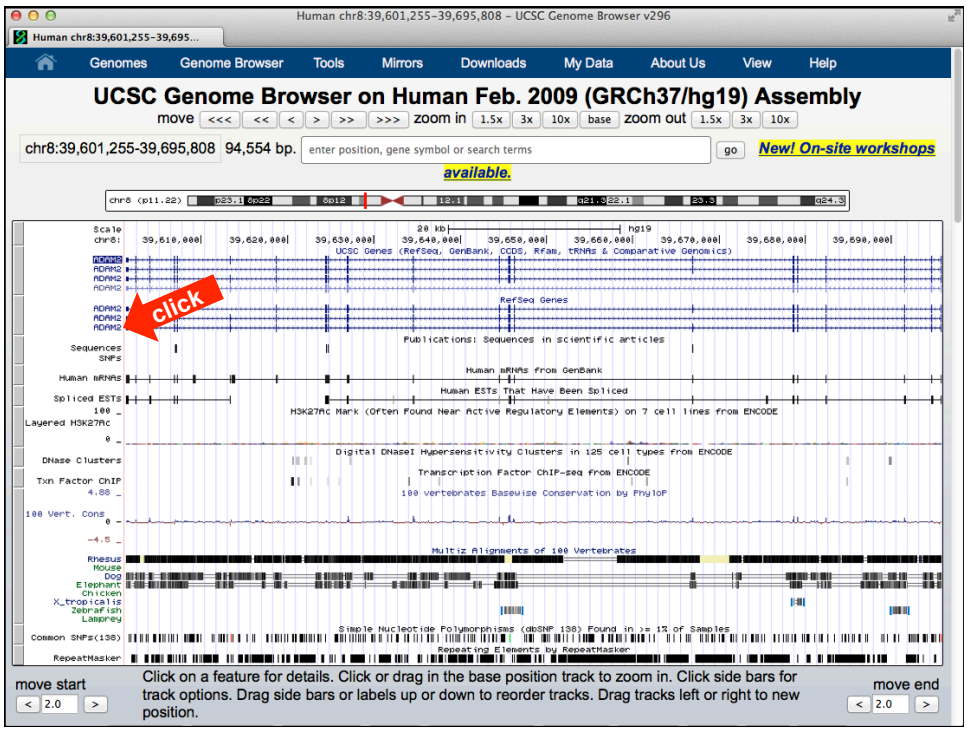


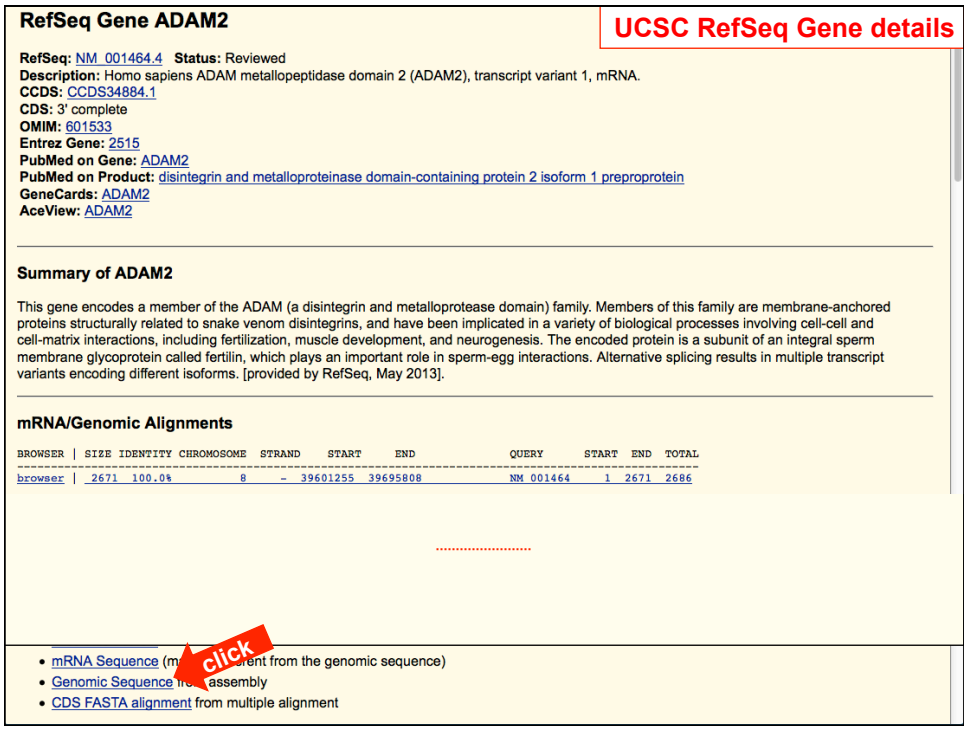

Human (Homo sapiens) Genome Browser Gateway
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
Help
Human (Homo sapiens) Genome Browser Gateway
group
genome
assembly
position
search term
Mammal
Human
Feb. 2009 (GRCh37/hg19)
chr8:39601255-39695808
adam2
Dec. 2013 (GRCh38/hg38)
Feb. 2009 (GRCh37/hg19)
Mar. 2006 (NCBI36/hg18)
May 2004 (NCBI35/hg17)
July 2003 (NCBI34/hg16)
ADAM2 (Homo sapiens ADAM metallopeptidase domain 2 (ADAM2), transcript variant 1, mRNA.)
ADAM20 (Homo sapiens ADAM metallopeptidase domain 20 (ADAM20), mRNA.)
ADAM20P1 (Homo sapiens ADAM metallopeptidase domain 20 pseudogene 1 (ADAM20P1), non-coding RNA.)
ADAM21 (Homo sapiens ADAM metallopeptidase domain 21 (ADAM21), mRNA.)
ADAM21P1 (Homo sapiens ADAM metallopeptidase domain 21 pseudogene 1 (ADAM21P1), non-coding RNA.)
ADAM22 (Homo sapiens ADAM metallopeptidase domain 22 (ADAM22), transcript variant 1, mRNA.)
ADAM23 (Homo sapiens ADAM metallopeptidase domain 23 (ADAM23), mRNA.)
ADAM28 (Homo sapiens ADAM metallopeptidase domain 28 (ADAM28), transcript variant 1, mRNA.)
ADAM29 (Homo sapiens ADAM metallopeptidase domain 29 (ADAM29), transcript variant 7, mRNA.)
Human Genome Browser – hg19 assembly (sequences)
Sample position queries
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

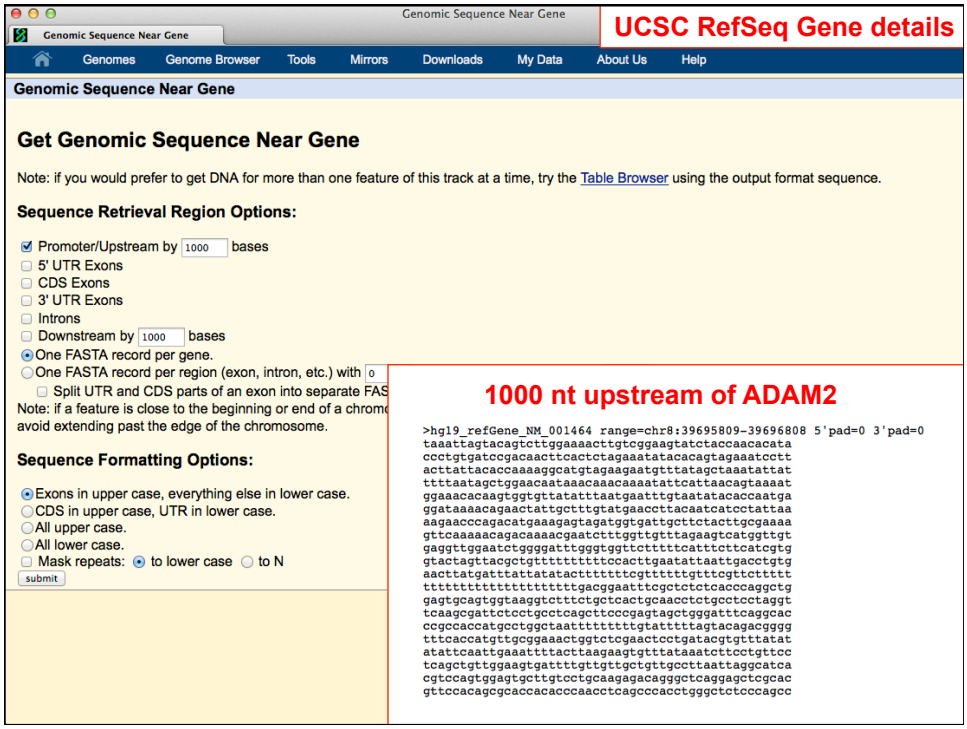

Human chr8:39601255-39695808 - UCSC Genome Browser v296
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
View
Help
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move
zoom in
zoom out
chr8:39,601,255-39,695,808
94,554 bp.
enter position, gene symbol or search terms
go
New! On-site workshops
available.
click
exon
intron
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
RefSeq Genes
Publications: Sequences in scientific articles
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE
Transcription Factor ChIP-seq from ENCODE
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples
Repeating Elements by RepeatMasker
move start
move end
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

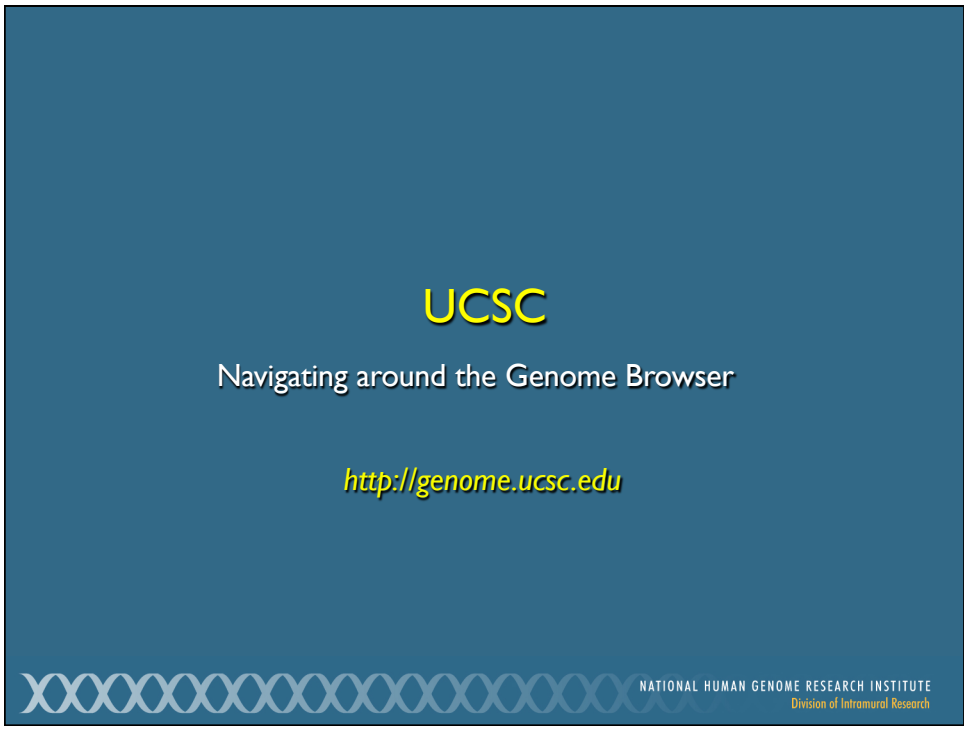

UCSC Gene details
Human Gene ADAM2 (uc003xnj.4) Description and Page Index
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
Help
Human Gene ADAM2 (uc003xnj.4) Description and Page Index
Description: Homo sapiens ADAM metallopeptidase domain 2 (ADAM2), transcript variant 1, mRNA.
RefSeq 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
Ratios
Absolute
Ratios
Absolute

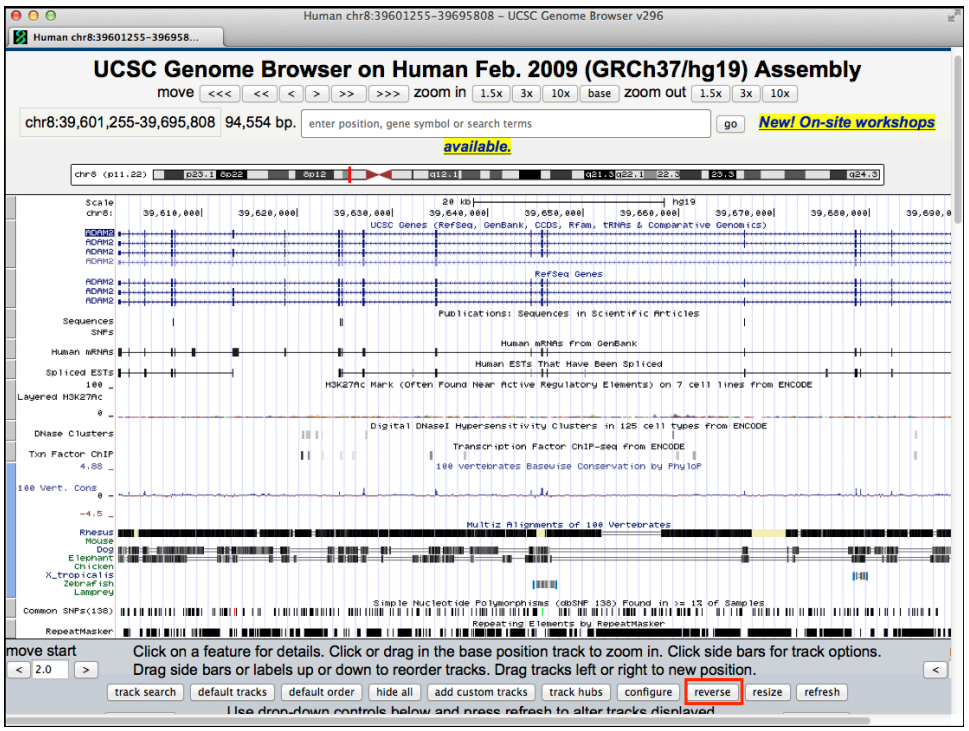

UCSC Change Track Display
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
RefSeq Genes
Publications: Sequences in scientific articles
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE
Transcription Factor ChIP-seq from ENCODE
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples
Repeating Elements by RepeatMasker
move start
move end
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
Genes and Gene Predictions
click
UCSC Genes
RefSeq Genes
AceView Genes
CCDS
Ensembl Genes
EvoFold
Exoniphy
Geneid Genes
Genscan Genes
H-Inv 7.0
IKMC Genes Mapped
hide
dense
squish
pack
full

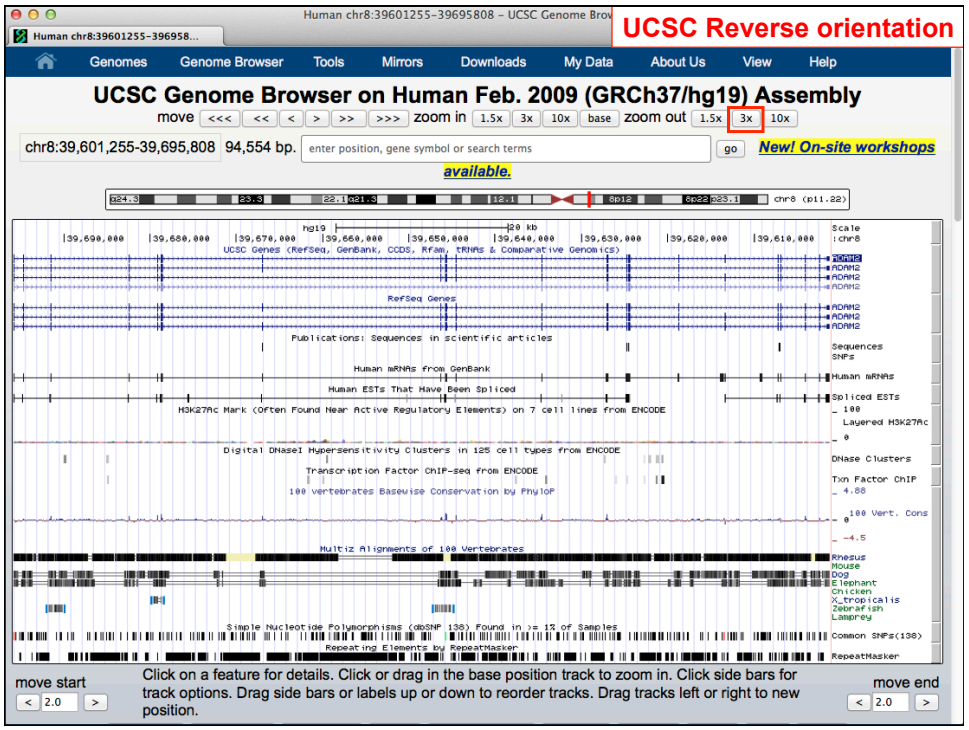

Human chr8:39,601,255-39,695,808 - UCSC Genome Browser v296
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
View
Help
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
chr8:39,601,255-39,695,808
94,554 bp.
go
New! On-site workshops available.
click
RefSeq Genes
Publications: Sequences in scientific articles
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE
Transcription Factor ChIP-seq from ENCODE
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples
Repeating Elements by RepeatMasker
move start
move end
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.

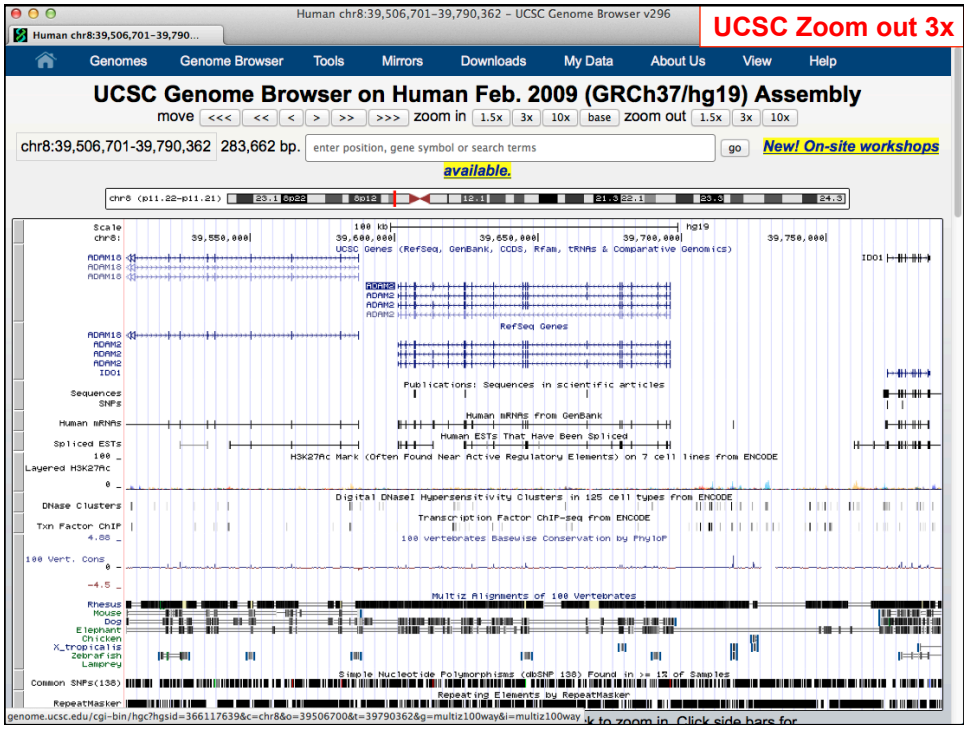

RefSeq Gene ADAM2
UCSC RefSeq Gene details
RefSeq: NM_001464.4 Status: Reviewed
Description: Homo sapiens ADAM metallopeptidase domain 2 (ADAM2), transcript variant 1, mRNA.
CCDS: CCDS34884.1
CDS: 3' complete
OMIM: 601533
Entrez Gene: 2515
PubMed on Gene: ADAM2
PubMed on Product: disintegrin and metalloproteinase 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
click
mRNA Sequence
Genomic Sequence
CDS FASTA alignment from multiple alignment

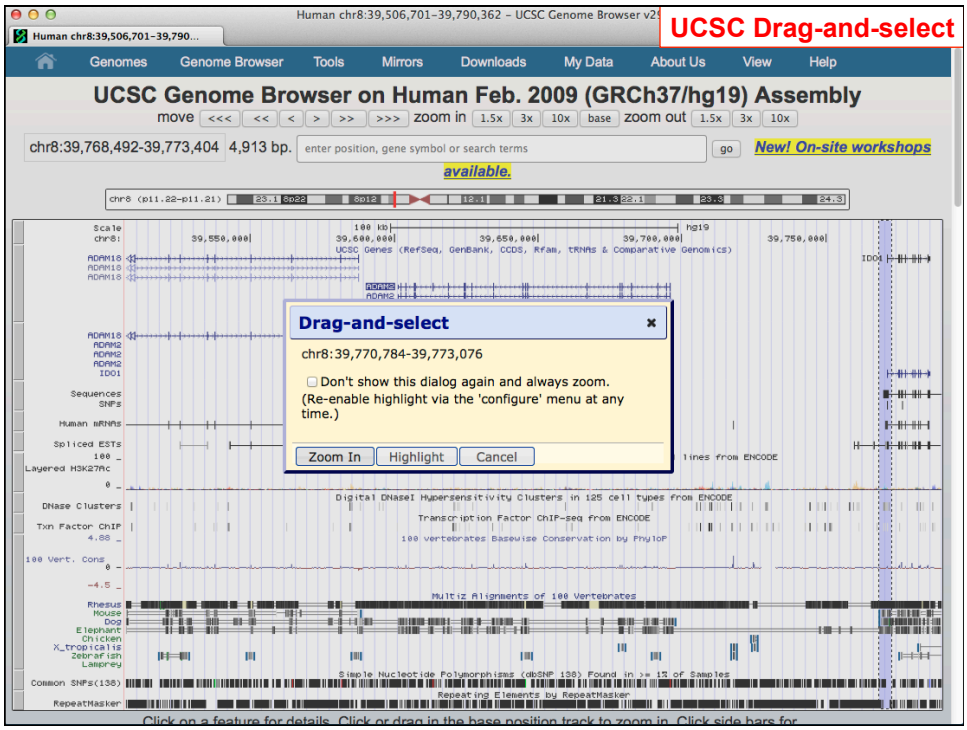

Genomic Sequence Near Gene
Genomic Sequence Near Gene
UCSC RefSeq Gene details
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
Help
Genomic Sequence Near Gene
Get Genomic Sequence Near Gene
Note: if you would prefer to get DNA for more than one feature of this track at a time, try the Table Browser using the output format sequence.
Sequence Retrieval Region Options:
Promoter/Upstream by 1000 bases
5' UTR Exons
CDS Exons
3' UTR Exons
Introns
Downstream by 1000 bases
One FASTA record per gene.
One FASTA record per region (exon, intron, etc.) with 0
Split UTR and CDS parts of an exon into separate FAS
Note: if a feature is close to the beginning or end of a chrom
avoid extending past the edge of the chromosome.
Sequence Formatting Options:
Exons in upper case, everything else in lower case.
CDS in upper case, UTR in lower case.
All upper case.
All lower case.
Mask repeats: to lower case to N
submit
1000 nt upstream of ADAM2
>hg19_refGene_NM_001464 range=chr8:39695809-39696808 5'pad=0 3'pad=0
taaattagtacagtcttggaaaacttgtcggaagtatctaccaacacata
ccctgtgatccgacaacttcactctagaaatatacacagtagaaatcctt
acttattacaccaaaaggcatgtagaagaatgtttatagctaaatattat
ttttaatagctggaacaataaacaaacaaaatattcattaacagtaaaat
ggaaacacaagtggtgttatatttaatgaatttgtaatatacaccaatga
ggataaaacagaactattgctttgtatgaaccttacaatcatcctattaa
aagaacccagacatgaaagagtagatggtgattgcttctacttgcgaaaa
gttcaaaaacagacaaaacgaatctttggttgtttagaagtcatggttgt
gaggttggaatctgggggatttgggtggttcttttctcatttcttcatcgtg
gtactagttacgctgtttttttttttccacttgaatattaattgacctgtg
aacttatgatttattatatactttttttcgtttttgtttcgttctttttt
ttttttttttttttttttttgacggaatttcgctctctcaccaggctg
gagtgcagtggtaaggtctttctgctcactgcaacctctgcctcctaggt
tcaagcgattctcctgcctcagcttcccgagtagctgggatttcaggcac
ccgccaccatgcctggctaatttttttttgtatttttagtacagacgggg
tttcaccatgttgcggaaactggtctcgaactcctgatacgtgtttatat
atattcaattgaaatttacttaagaagtgtttataaatcttcctgttcc
tcagctgttggaagtgattttgttgttgctgttgccttaattaggcatca
cgtccagtggagtgcttgtcctgcaagagacagggctcaggagctcgcac
gttccacagcgcaccacacccaacctcagcccacctgggctctcccagcc

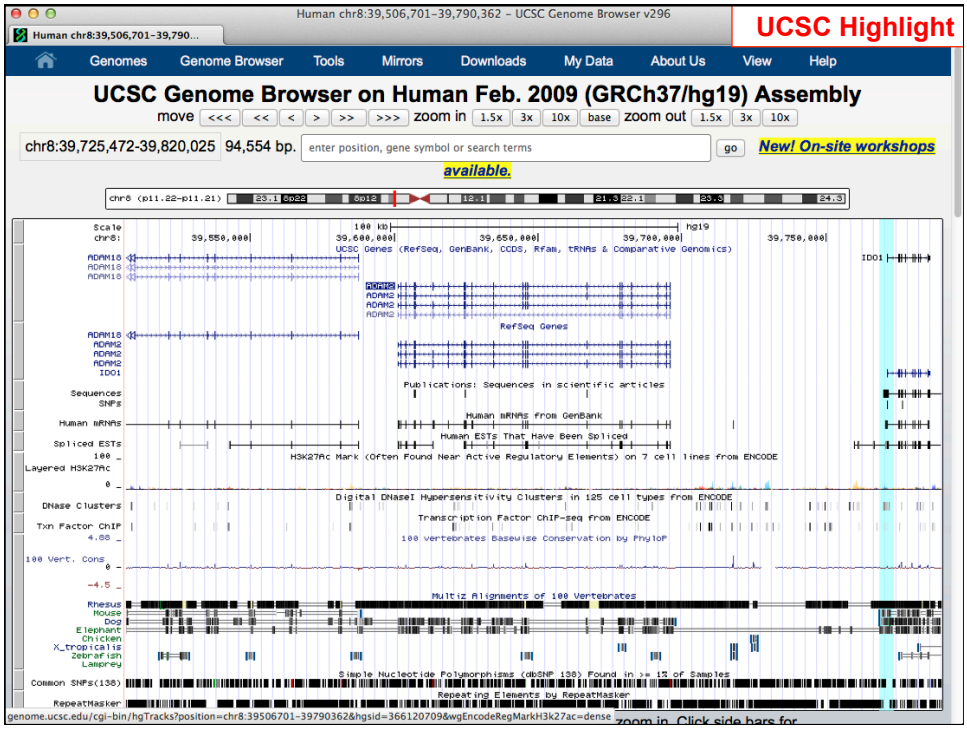

UCSC
Navigating around the Genome Browser
http://genome.ucsc.edu
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

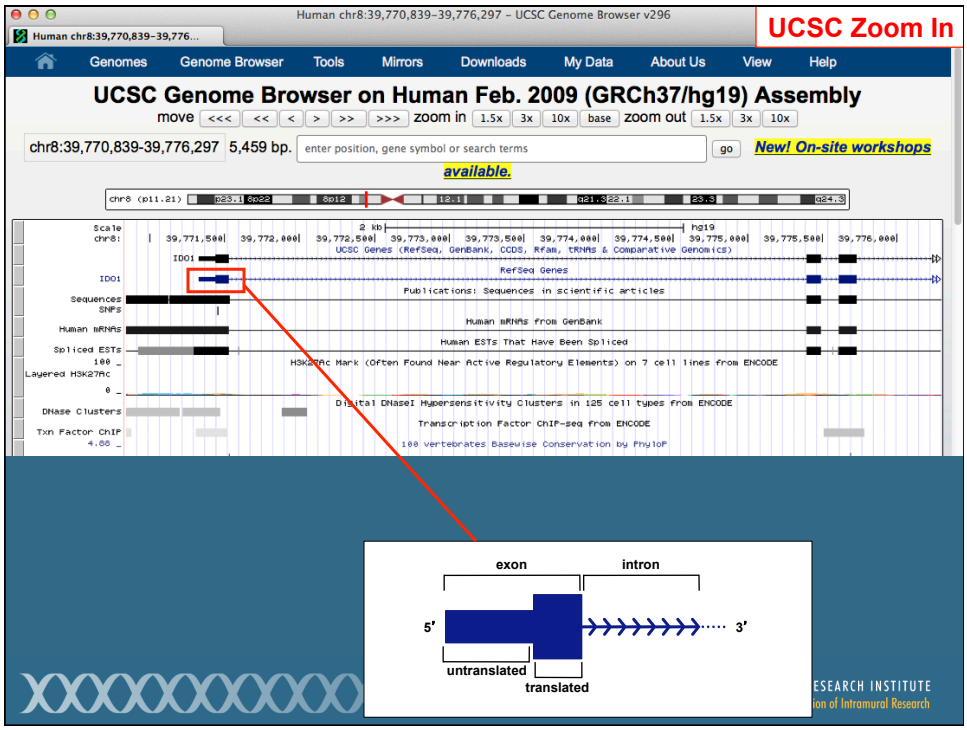

Human chr8:39601255-39695808 - UCSC Genome Browser v296
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move zoom in zoom out
chr8:39,601,255-39,695,808 94,554 bp.
enter position, gene symbol or search terms
go
New! On-site workshops available.
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
RefSeq Genes
Publications: Sequences in Scientific Articles
Human mRNAs from GenBank
Human ESTs That Have Been Spliced
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE
Transcription Factor ChIP-seq from ENCODE
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples
Repeating Elements by RepeatMasker
move start
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


UCSC Reverse orientation
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
View
Help
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
chr8:39,601,255-39,695,808 94,554 bp.
New! On-site workshops available.
move start
move end
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.


UCSC Zoom out 3x
Human chr8:39,506,701-39,790,362 – UCSC Genome Browser v296
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
View
Help
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
chr8:39,506,701-39,790,362
283,662 bp.
New! On-site workshops
available.


UCSC Drag-and-select
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
View
Help
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
chr8:39,768,492-39,773,404
4,913 bp.
New! On-site workshops
available.
Drag-and-select
chr8:39,770,784-39,773,076
Don't show this dialog again and always zoom.
(Re-enable highlight via the 'configure' menu at any time.)
Zoom In
Highlight
Cancel


UCSC Highlight
Human chr8:39,506,701-39,790,362 - UCSC Genome Browser v296
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
View
Help
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
chr8:39,725,472-39,820,025
94,554 bp.
New! On-site workshops
available.

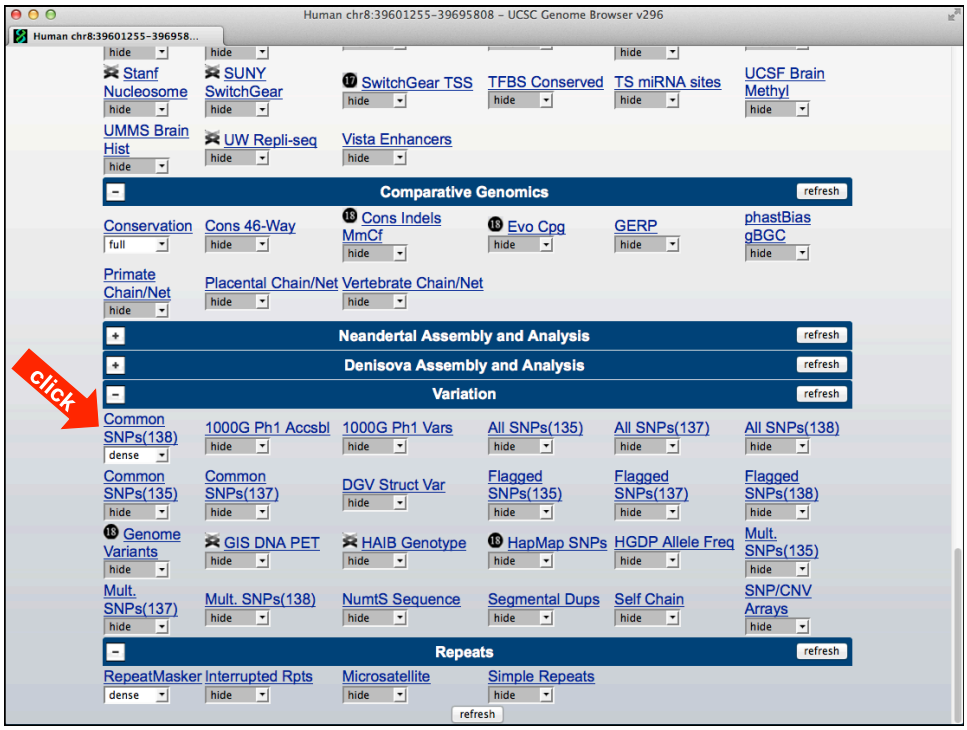

UCSC Zoom In
Human chr8:39,770,839-39,776,297 - UCSC Genome Browser v296
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
View
Help
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
chr8:39,770,839-39,776,297
5,459 bp.
New! On-site workshops
available.
exon
intron
5'
3'
untranslated
translated

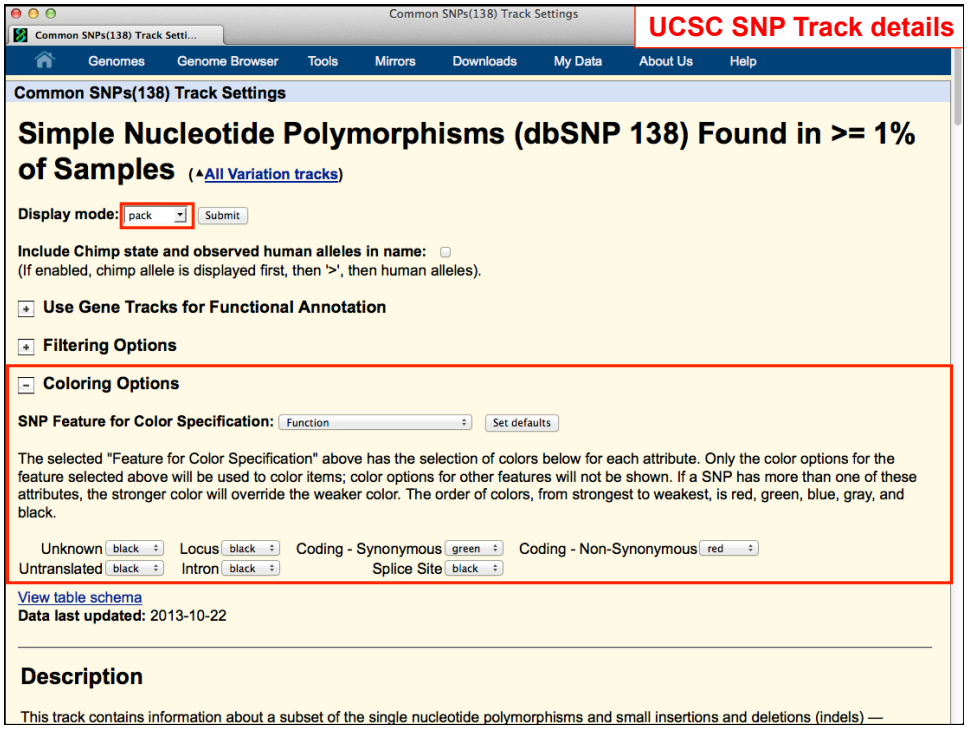
UCSC

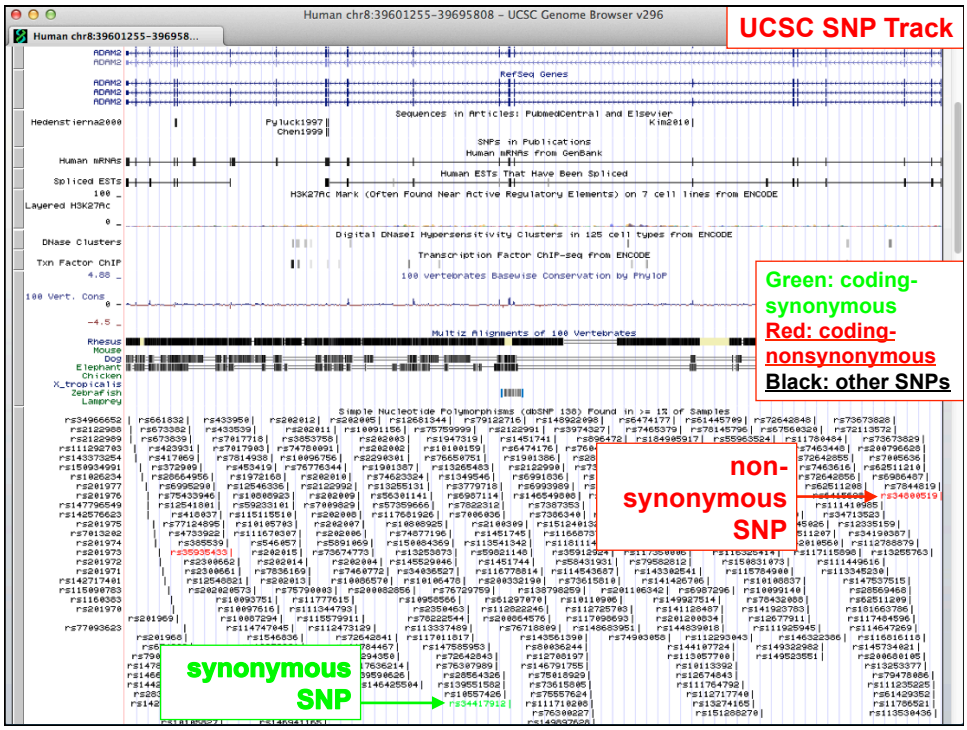
Configure Track on the Genome Browser

http://genome.ucsc.edu

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research




UCSC SNP Track details
Common SNPs(138) Track Settings
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) —

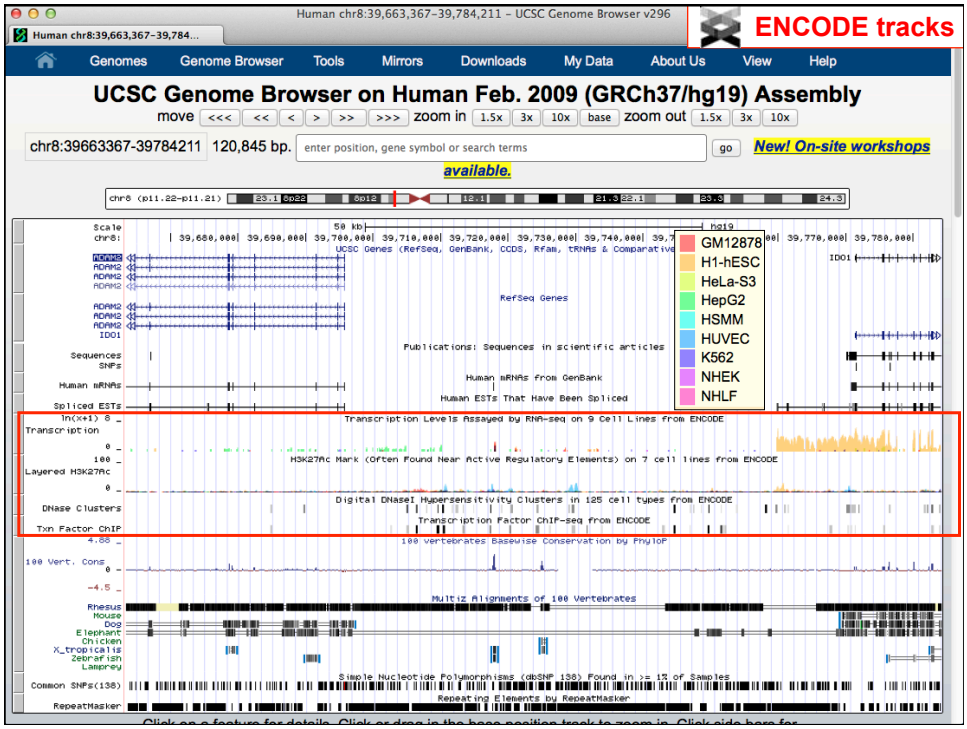

UCSC SNP Track
Human chr8:39601255-39695808 - UCSC Genome Browser v296
Green: coding-synonymous
Red: coding-nonsynonymous
Black: other SNPs
non-synonymous SNP
synonymous SNP

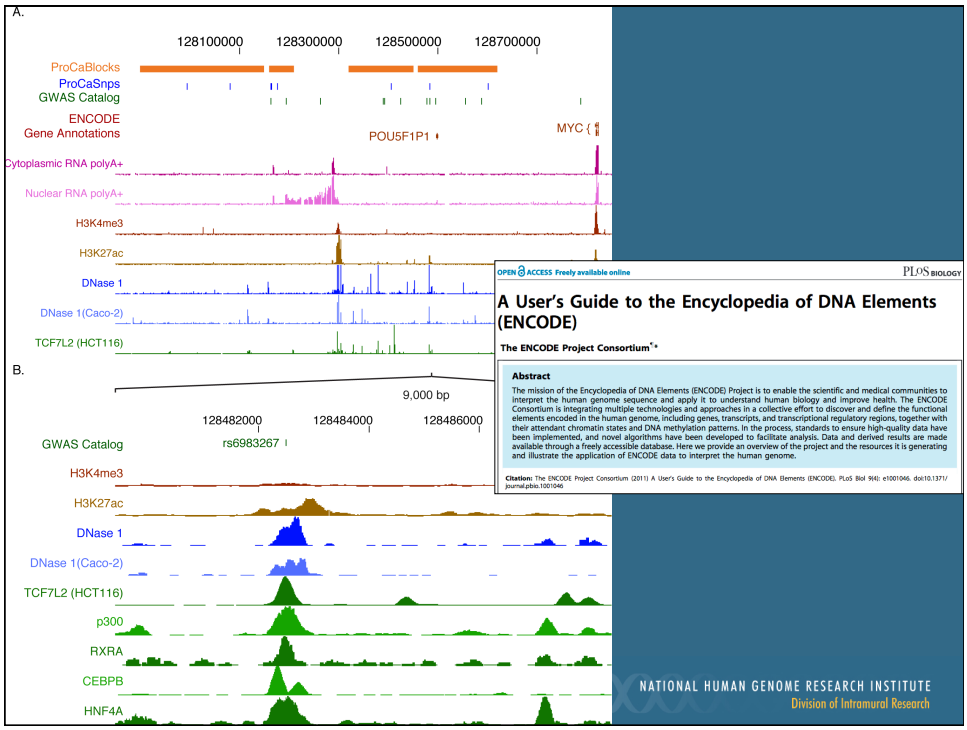
UCSC

ENCODE tracks

http://genome.ucsc.edu

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research


Human chr8:39,663,367-39,784,211 - UCSC Genome Browser v296
ENCODE tracks
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
View
Help
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
chr8:39663367-39784211 120,845 bp.
New! On-site workshops available.
GM12878
H1-hESC
HeLa-S3
HepG2
HSMM
HUVEC
K562
NHEK
NHLF
RefSeq Genes
Transcription Levels Assayed by RNA-seq on 9 Cell Lines from ENCODE
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE
Transcription Factor ChIP-seq from ENCODE
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples
Repeating Elements by RepeatMasker

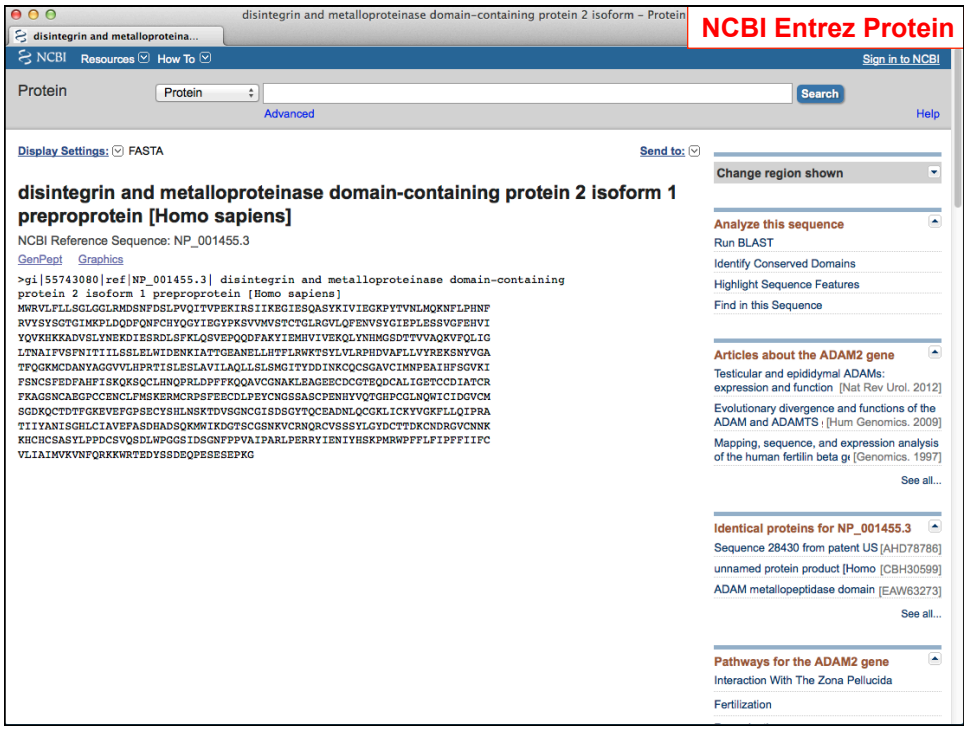

A.
128100000
128300000
128500000
128700000
ProCaBlocks
ProCaSnps
GWAS Catalog
ENCODE Gene Annotations
POU5F1P1
MYC
Cytoplasmic RNA polyA+
Nuclear RNA polyA+
H3K4me3
H3K27ac
DNase 1
DNase 1(Caco-2)
TCF7L2 (HCT116)
B.
9,000 bp
128482000
128484000
128486000
GWAS Catalog
rs6983267
H3K4me3
H3K27ac
DNase 1
DNase 1(Caco-2)
TCF7L2 (HCT116)
p300
RXRA
CEBPB
HNF4A
OPEN ACCESS Freely available online
PLoS BIOLOGY
A User's Guide to the Encyclopedia of DNA Elements (ENCODE)
The ENCODE Project Consortium*
Abstract
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

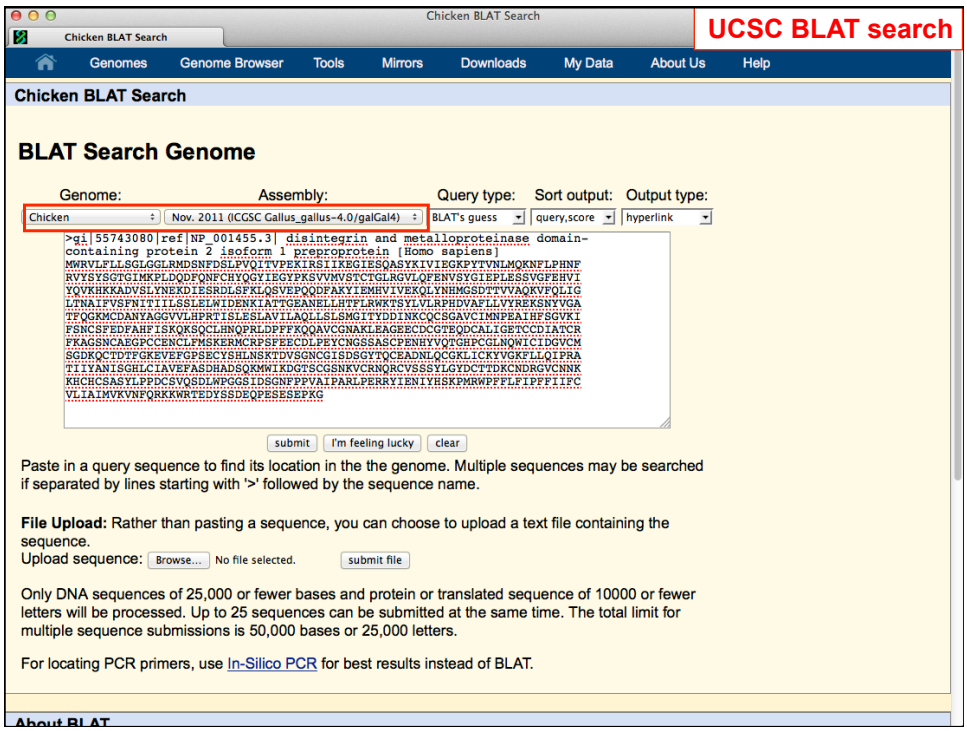
UCSC

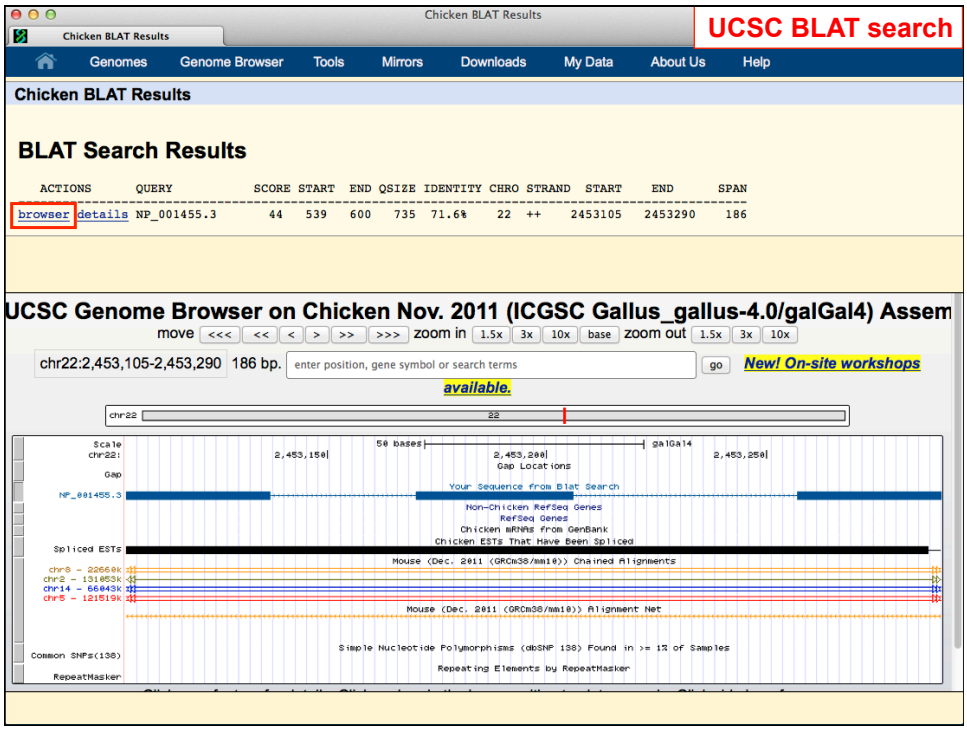
Find a chicken homolog of a human protein

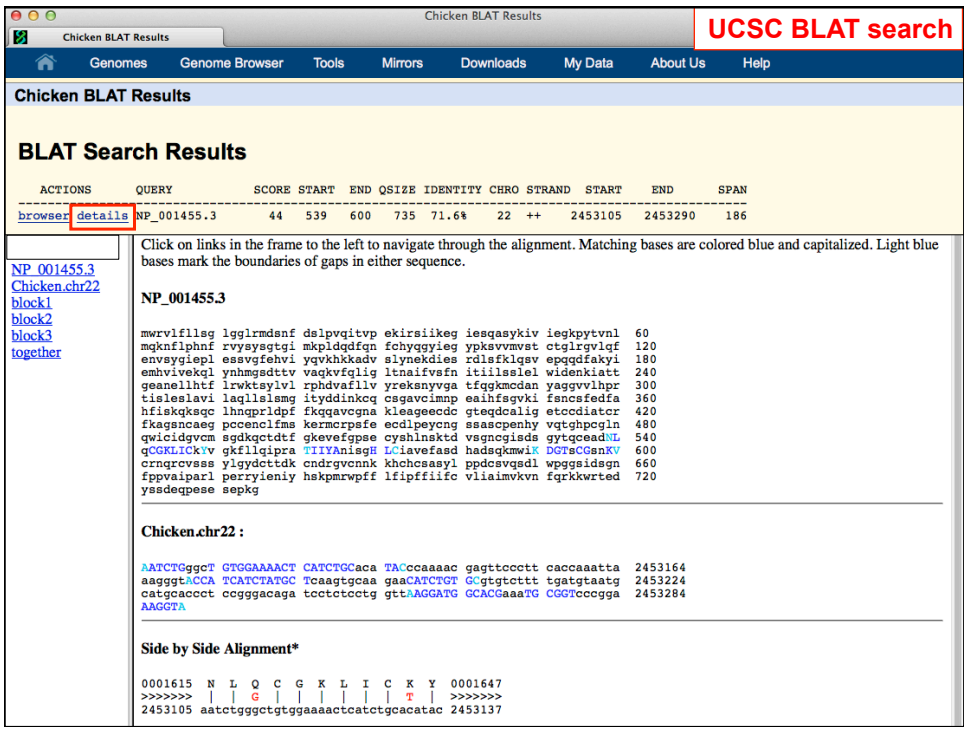
http://genome.ucsc.edu

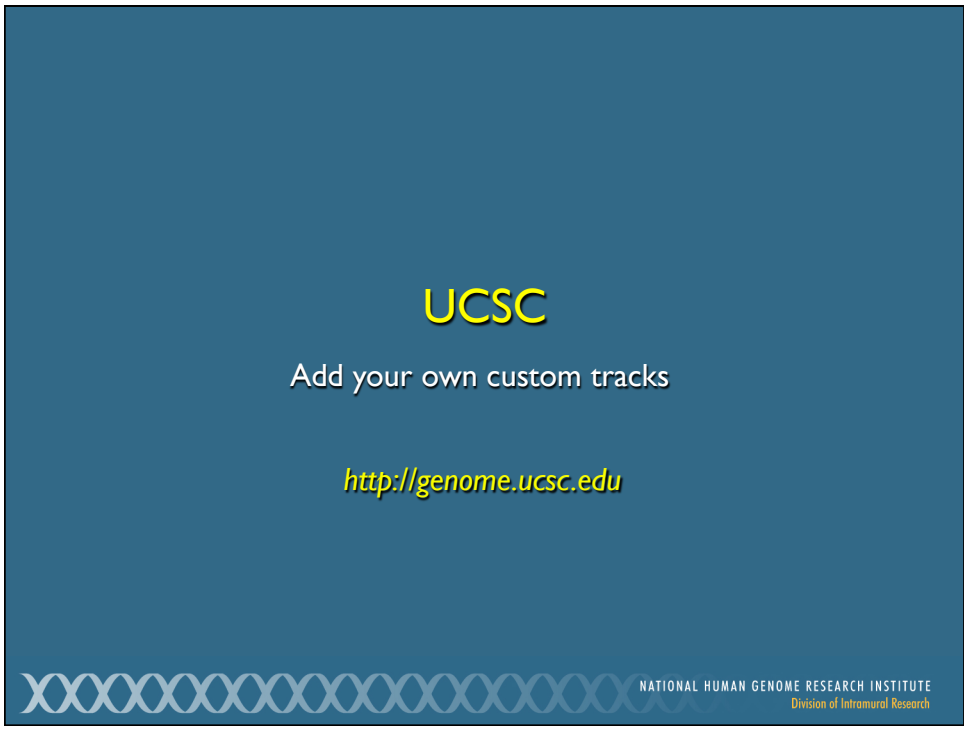
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research




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:
Chicken
Nov. 2011 (ICGSC Gallus_gallus-4.0/galGal4)
BLAT's guess
query,score
hyperlink
>gi|55743080|ref|NP_001455.3| disintegrin and metalloproteinase domain-containing protein 2 isoform 1 preproprotein [Homo sapiens]
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKIVIEGKPYTVNLMQKNFLPHNF
RVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPLESSVGFEHVI
YQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDPAKYIEMHVIVEKQLYNHMGSDTTVVAQKVFQLIG
LTNAIFVSFNITIILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVLRPHDVAFLLVYREKSNYVGA
TFQGKMCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAIHFSGVKI
FSNCSFEDFAHFISKQKSQCLHNQPRLDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCM
SGDKQCTDTFGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYTQCEADNLQCGKLICKYVGKFLLQIPRA
TIIYANISGHLCIAVEFASDHADSQKMWIKDGTSCGSNKVCRNQRCVSSSYLGYDCTTDKCNDRGVCNNK
KHCHCSASYLPPDCSVQSDLWPGGSIDSGNFPPVAIPARLPERRYIENIYHSKPMRWFFFLFIPFFIIFC
VLIAIMVKVNFQRKKWNTEDYSSDEQPESESEPKG
submit
I'm feeling lucky
clear
Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.
File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.
Upload sequence:
Browse...
No file selected.
submit file
Only DNA sequences of 25,000 or fewer bases and protein or translated sequence of 10000 or fewer letters will be processed. Up to 25 sequences can be submitted at the same time. The total limit for multiple sequence submissions is 50,000 bases or 25,000 letters.
For locating PCR primers, use In-Silico PCR for best results instead of BLAT.


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
zoom out
chr22:2,453,105-2,453,290 186 bp.
enter position, gene symbol or search terms
go
New! On-site workshops available.
Gap
Your Sequence from Blat Search
NP_001455.3
Non-Chicken RefSeq Genes
RefSeq Genes
Chicken mRNAs from GenBank
Chicken ESTs That Have Been Spliced
Spliced ESTs
Mouse (Dec. 2011 (GRCm38/mm10)) Chained Alignments
Mouse (Dec. 2011 (GRCm38/mm10)) Alignment Net
Simple Nucleotide Polymorphisms (dbSNP 138) Found in >= 1% of Samples
Repeating Elements by RepeatMasker
Common SNPs(138)
RepeatMasker


Chicken BLAT Results
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
NP_001455.3
Chicken.chr22
block1
block2
block3
together
Click on links in the frame to the left to navigate through the alignment. Matching bases are colored blue and capitalized. Light blue bases mark the boundaries of gaps in either sequence.
NP_001455.3
mwrvlfllsg lgglrmdsnf dslpvqitvp ekirsiikeg iesqasykiv iegkpytvnl 60
mqknflphnf rvysysgtgi mkpldqdfqn fchyqgyieg ypksvvmvst ctglrgvlqf 120
envsygiepl essvgfehvi yqvkhkkadv slynekdies rdlsfklqsv epqqdfakyi 180
emhvivekql ynhmgsdttv vaqkvfqlig ltnaifvsfn itiilsslel widenkiatt 240
geanellhtf lrwktsylvl rphdvafllv yreksnyvga tfqgkmcdan yaggvvlhpr 300
tisleslavi laqllslsmg ityddinkcq csgavcimnp eaihfsgvki fsncsfedfa 360
hfiskqksqc lhnqprldpf fkqqavcgna kleageecdc gteqdcalig etccdiatcr 420
fkagsncaeg pccenclfms kermcrpsfe ecdlpeycng ssascpenhy vqtghpcgln 480
qwicidgvcm sgdkqctdtf gkevefgpse cyshlnsktd vsgncgisds gytgceadNL 540
qCGKLICKYv gkfllqiprs TIIYAnisgH LCiavefasd hadsqkmwiK DGTsCGsnKV 600
crnqrcvsss ylgydcttdk cndrgvcnnk khchcsasyl ppdcsvqsdl wpggsidsgn 660
fppvaiparl perryieniy hskpmrwpff lfipffiifc vliaimvkvn fqrkkwrted 720
yssdeqpese sepkg
Chicken.chr22 :
AATCTGggcT GTGGAAAACT CATCTGCaca TACccaaaac gagttccctt caccaaatta 2453164
aagggtACCA TCATCTATGC Tcaagtgcaa gaaCATCTGT GCgtgtcttt tgatgtaatg 2453224
catgcaccct ccgggacaga tcctctcctg gttAAGGATG GCACGaaaTG CGGTcccgga 2453284
AAGGTA
Side by Side Alignment*
0001615 N L Q C G K L I C K Y 0001647
>>>>>>> | | G | | | | | T | >>>>>>>
2453105 aatctgggctgtggaaaactcatctgcacatac 2453137

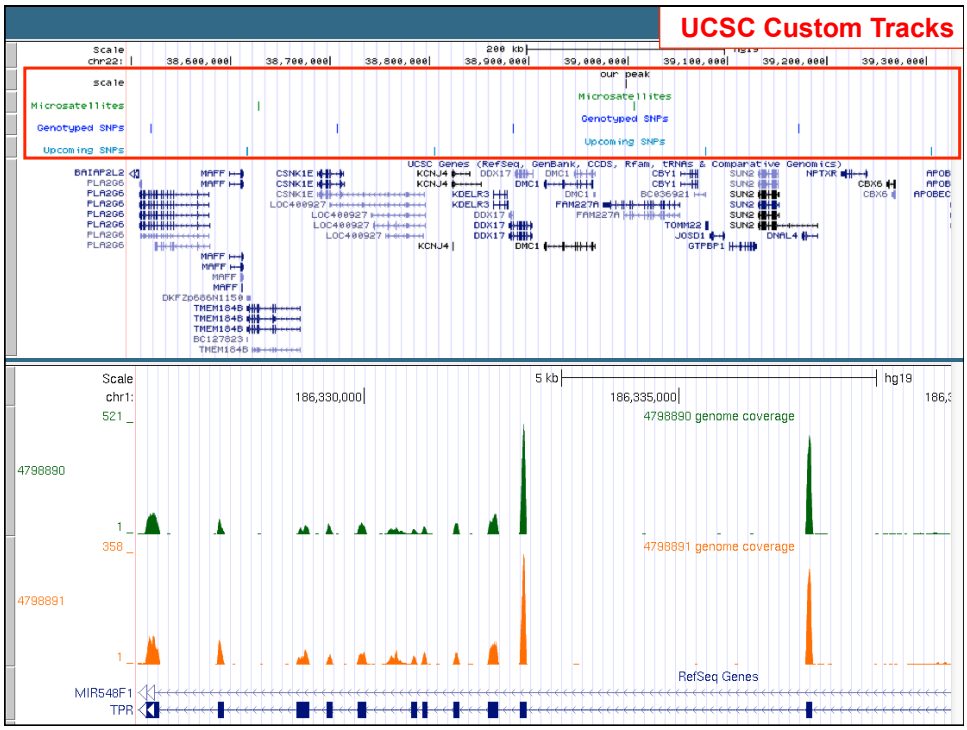

UCSC
Add your own custom tracks
http://genome.ucsc.edu
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

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

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research



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

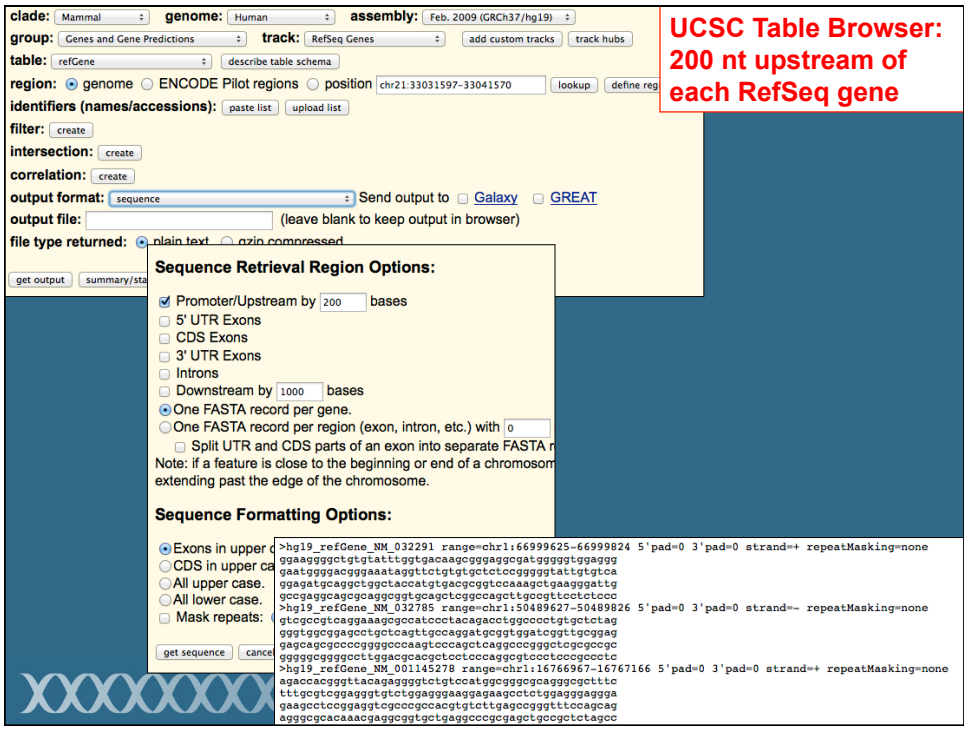
http://genome.ucsc.edu

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

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research



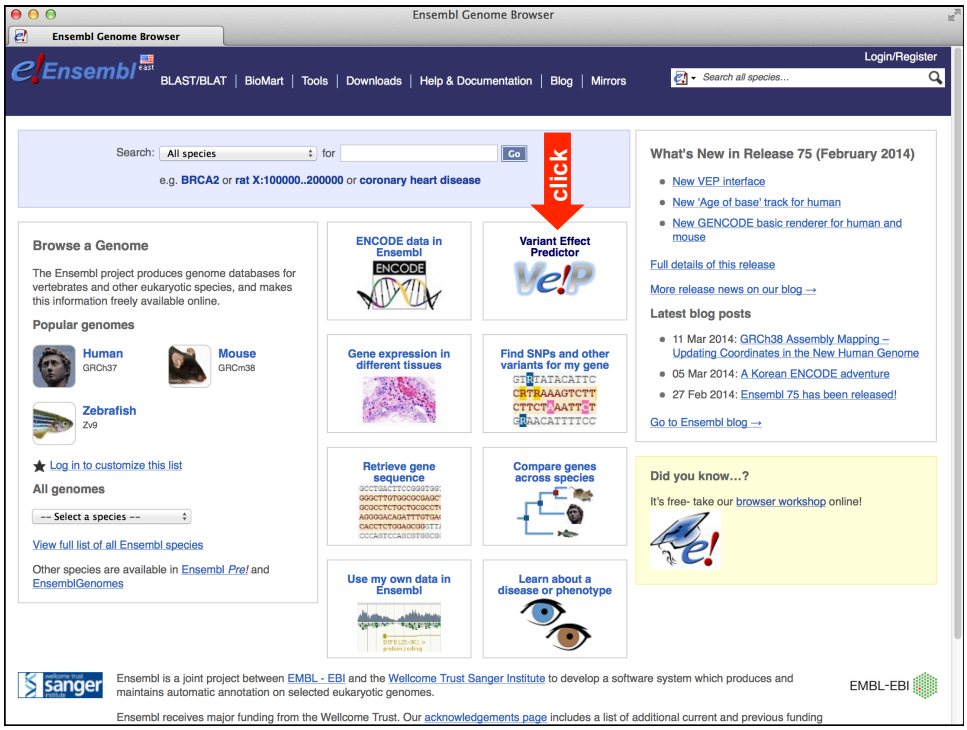
UCSC Table Browser: 200 nt upstream of each RefSeq gene

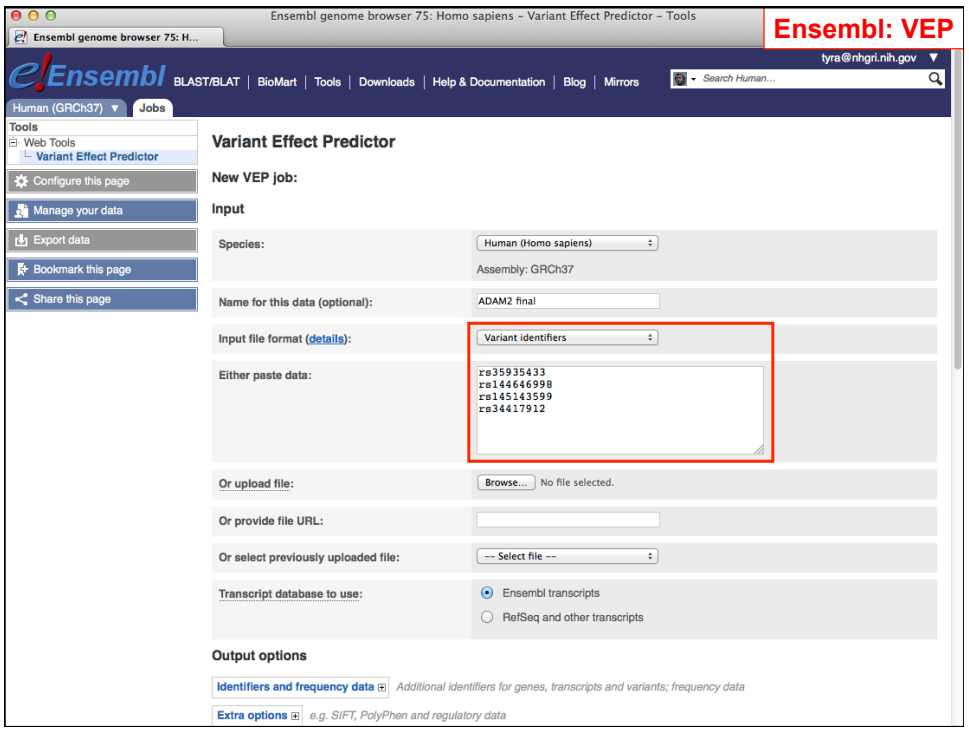
Ensembl

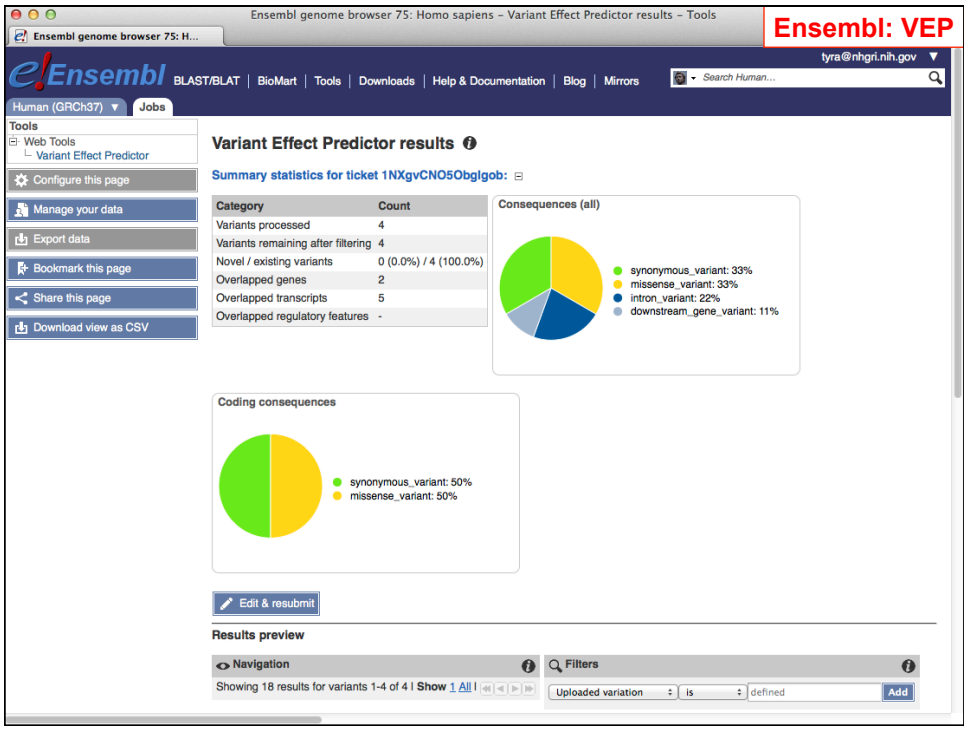
Variant Effect Predictor (VEP)

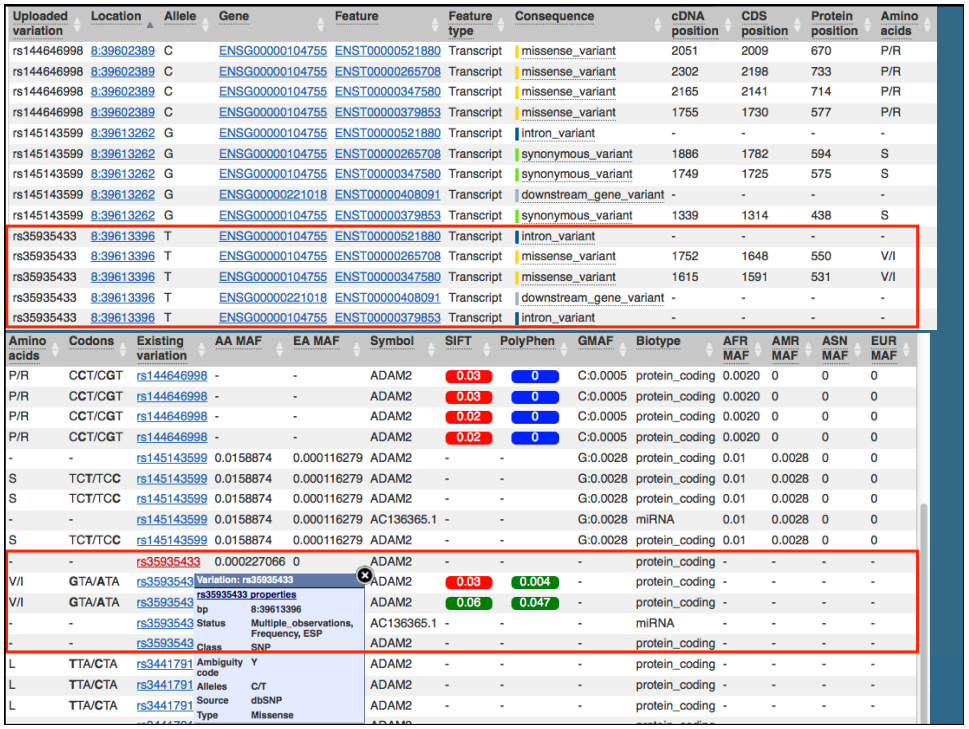
http://www.ensembl.org

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research



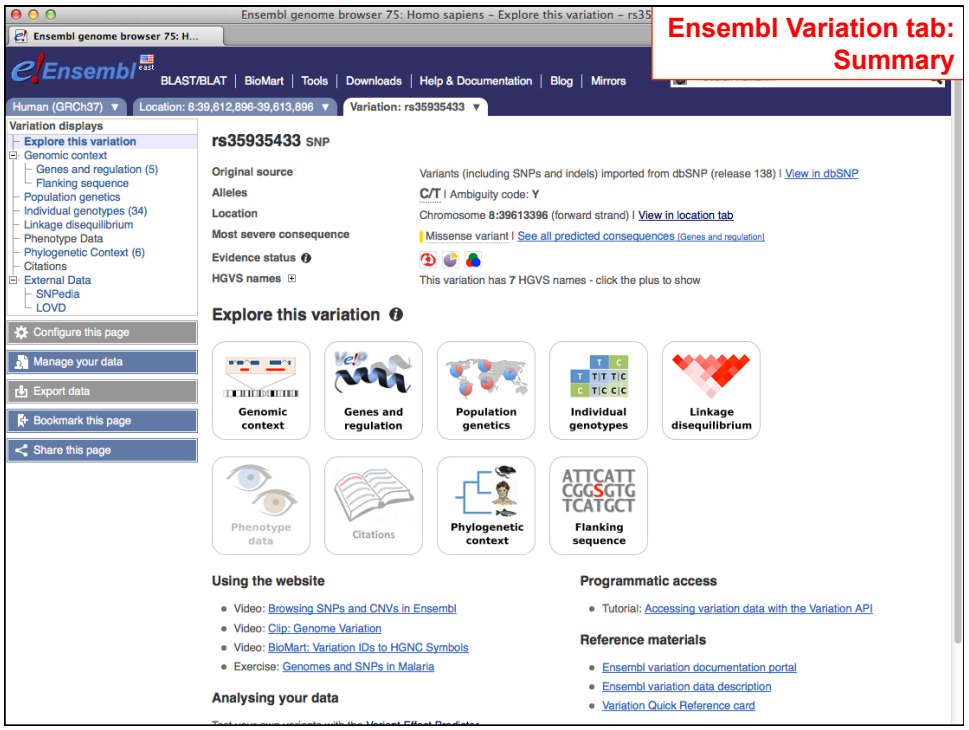

Ensembl: VEP
Ensembl genome browser 75: Homo sapiens - Variant Effect Predictor - Tools
tyra@nhgri.nih.gov
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Search Human...
Human (GRCh37) | Jobs
Tools
Web Tools
Variant Effect Predictor
Configure this page
Manage your data
Export data
Bookmark this page
Share this page
Variant Effect Predictor
New VEP job:
Input
Species:
Human (Homo sapiens)
Assembly: GRCh37
Name for this data (optional):
ADAM2 final
Input file format (details):
Variant identifiers
Either paste data:
rs35935433
rs144646998
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


Ensembl: VEP
Ensembl genome browser 75: Homo sapiens - Variant Effect Predictor results - Tools
tyra@nhgri.nih.gov
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Search Human...
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 1NXgvCNO5Obglgob:
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%
Edit & resubmit
Results preview
Navigation
Showing 18 results for variants 1-4 of 4 | Show 1 All |
Filters
Uploaded variation | is | defined | Add

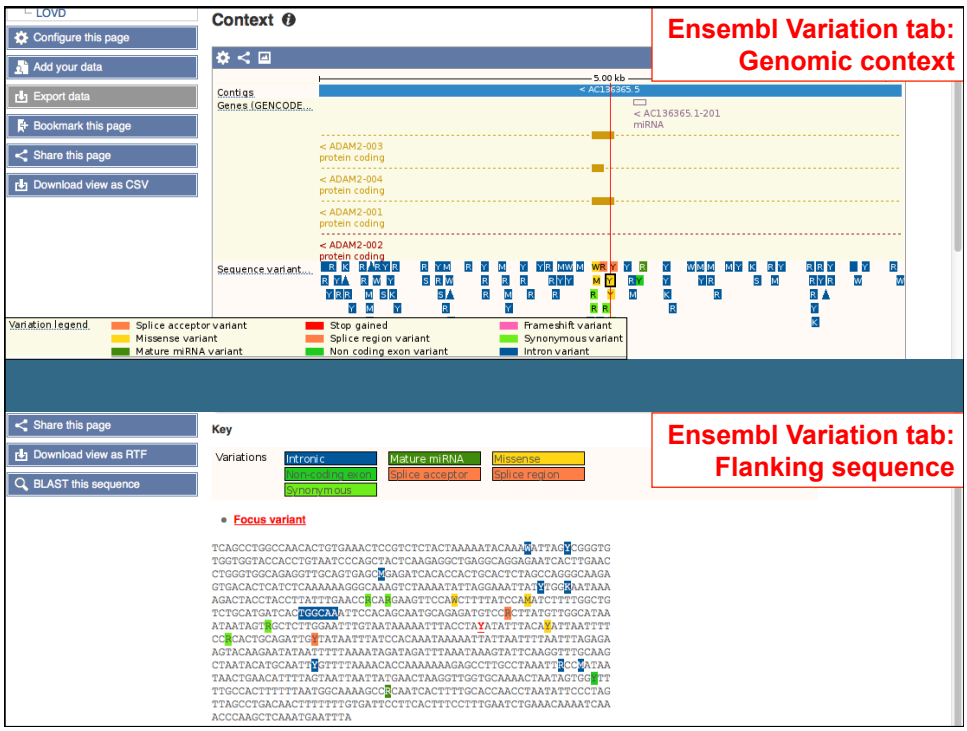


Uploaded variation	Location	Allele	Gene	Feature	Feature type	Consequence	cDNA position	CDS position	Protein position	Amino acids
rs144646998	8:39602389	C	ENSG00000104755	ENST00000521880	Transcript	missense_variant	2051	2009	670	P/R
rs144646998	8:39602389	C	ENSG00000104755	ENST00000265708	Transcript	missense_variant	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	-	-	-	-

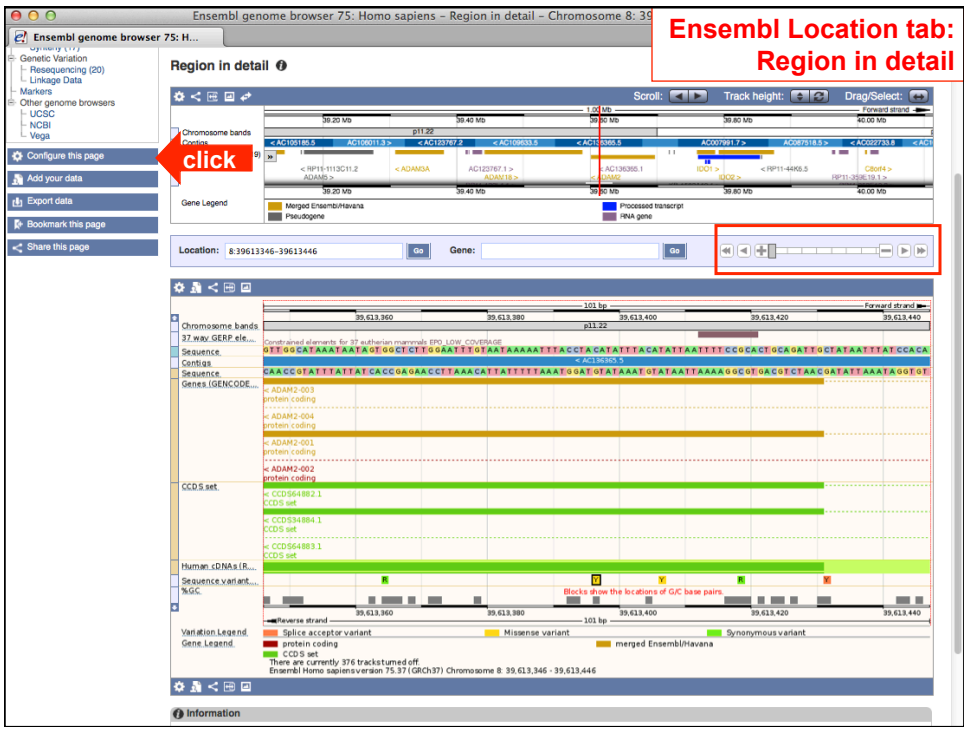
Amino acids	Codons	Existing variation	AA MAF	EA MAF	Symbol	SIFT	PolyPhen	GMAF	Biotype	AFR MAF	AMR MAF	ASN MAF	EUR MAF
P/R	CCT/CGT	rs144646998	-	-	ADAM2	0.03	0	C:0.0005	protein_coding	0.0020	0	0	0
P/R	CCT/CGT	rs144646998	-	-	ADAM2	0.03	0	C:0.0005	protein_coding	0.0020	0	0	0
P/R	CCT/CGT	rs144646998	-	-	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	rs3593543	Variation: rs35935433		ADAM2	0.03	0.004	-	protein_coding	-	-	-	-
V/I	GTA/ATA	rs3593543	rs35935433 properties		ADAM2	0.06	0.047	-	protein_coding	-	-	-	-
-	-	rs3593543	Status: Multiple_observations, Frequency, ESP		AC136365.1	-	-	-	miRNA	-	-	-	-
-	-	rs3593543	Class: SNP		ADAM2	-	-	-	protein_coding	-	-	-	-
L	TTA/CTA	rs3441791	Ambiguity code: Y		ADAM2	-	-	-	protein_coding	-	-	-	-
L	TTA/CTA	rs3441791	Alleles: C/T		ADAM2	-	-	-	protein_coding	-	-	-	-
L	TTA/CTA	rs3441791	Source: dbSNP		ADAM2	-	-	-	protein_coding	-	-	-	-
			Type: Missense										

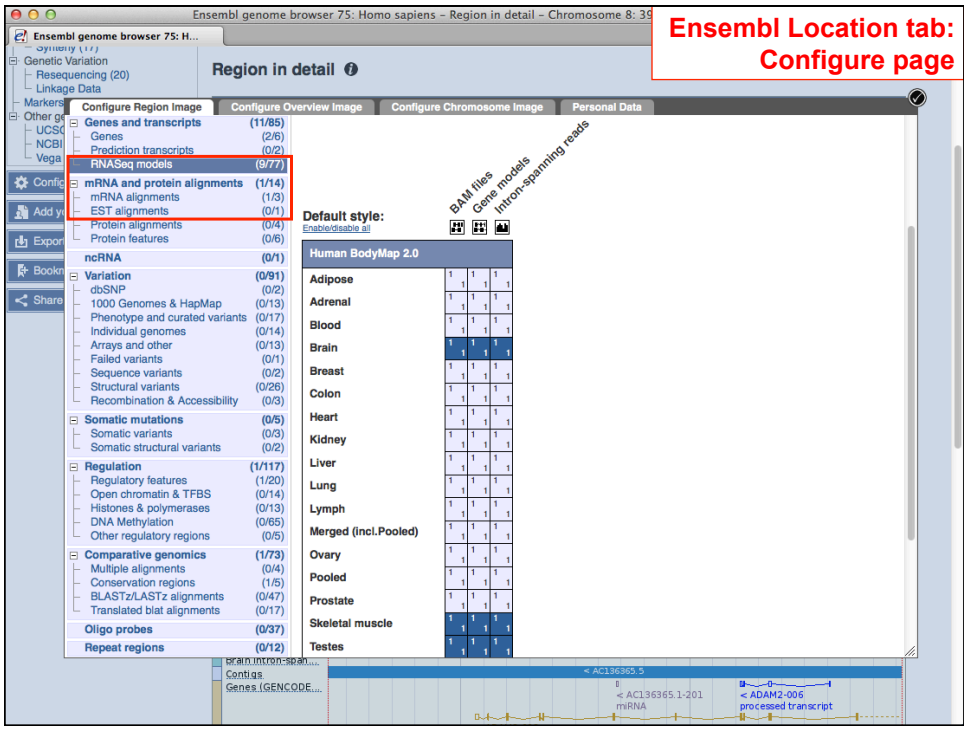


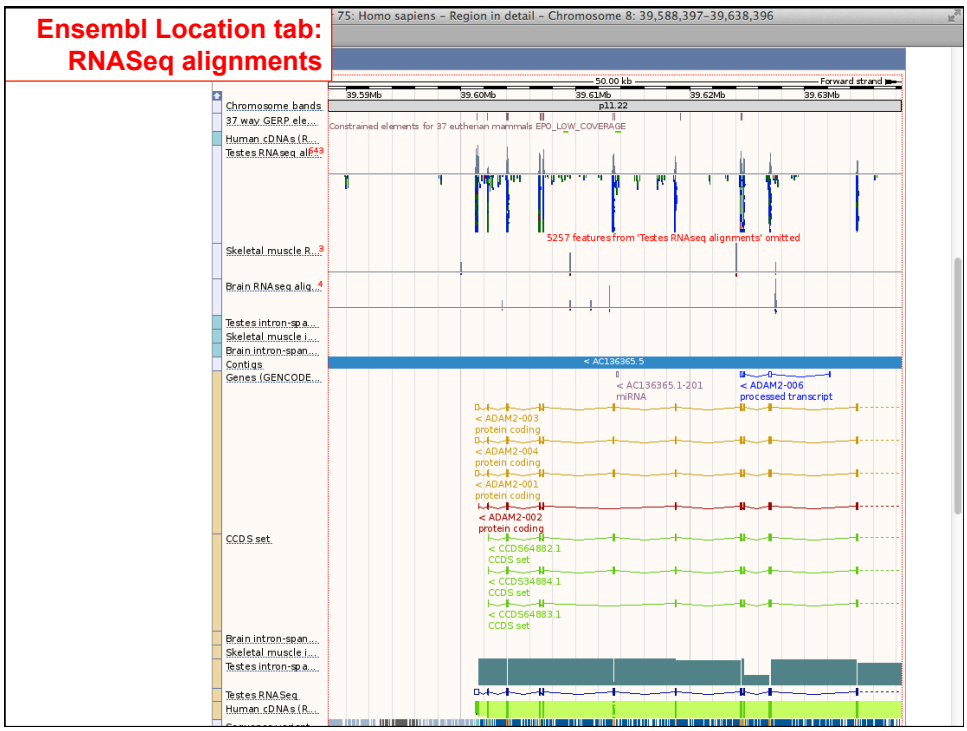
Ensembl Variation tab: Summary

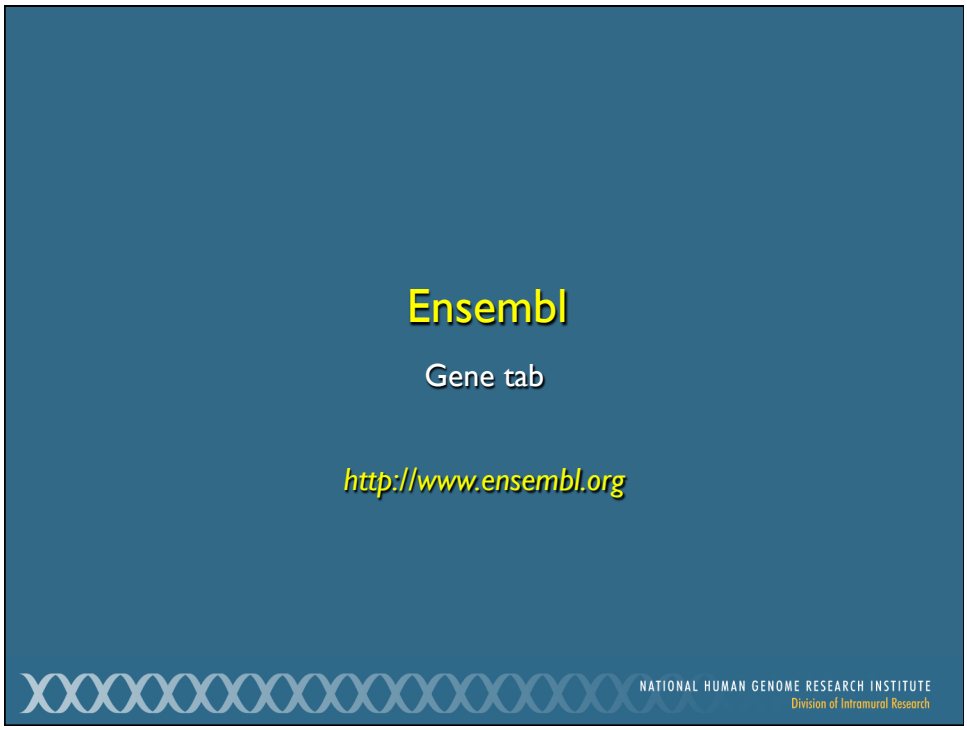

Ensembl Variation tab:
Genomic context
Context
Configure this page
Add your data
Export data
Bookmark this page
Share this page
Download view as CSV
Contigs
Genes (GENCODE...
Sequence variant...
Variation legend
Splice acceptor variant
Missense variant
Mature miRNA variant
Stop gained
Splice region variant
Non coding exon variant
Frameshift variant
Synonymous variant
Intron variant
Ensembl Variation tab:
Flanking sequence
Share this page
Download view as RTF
BLAST this sequence
Key
Variations
Intronic
Mature miRNA
Missense
Splice acceptor
Splice region
Synonymous
Focus variant

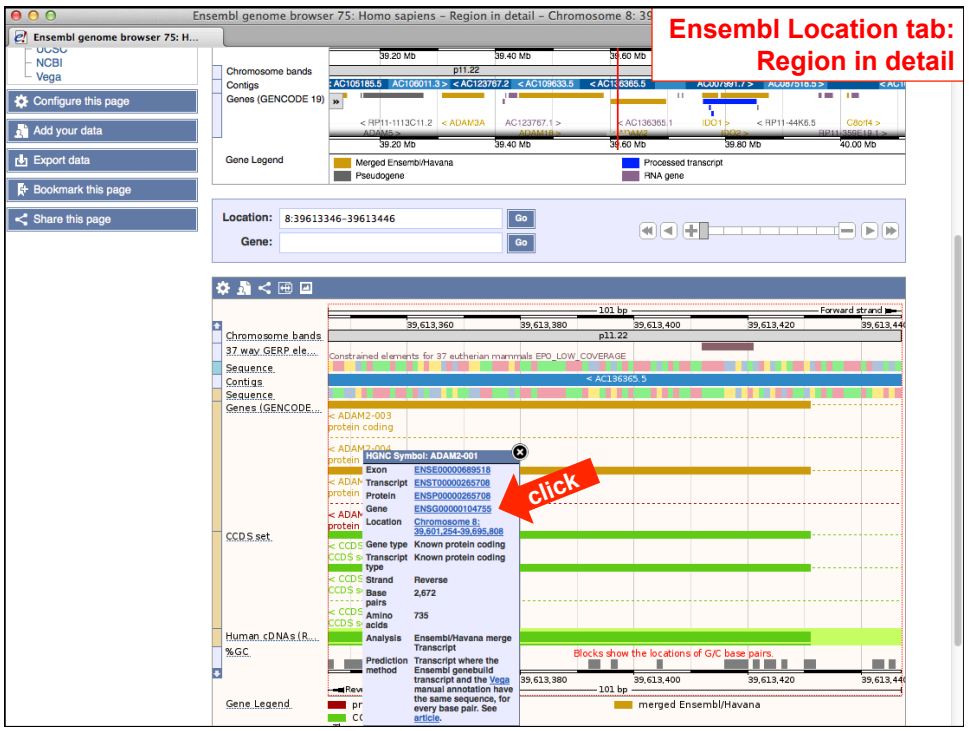

Ensembl
Location tab
http://www.ensembl.org
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

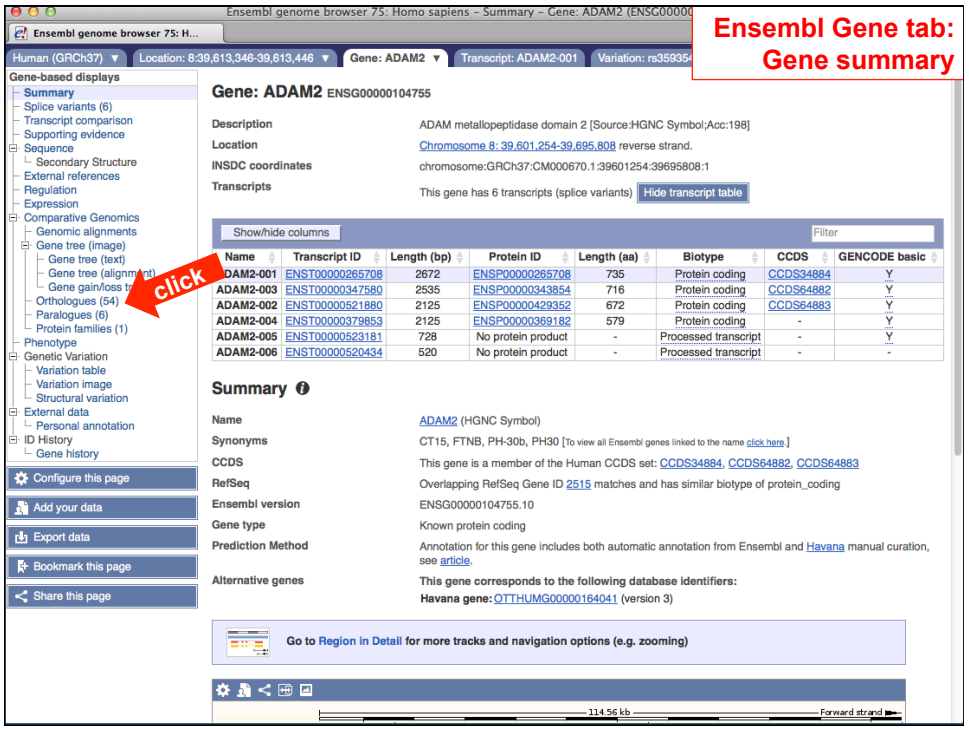

Ensembl Location tab:
Region in detail
Region in detail
click
Configure this page
Add your data
Export data
Bookmark this page
Share this page
Location:
Gene:

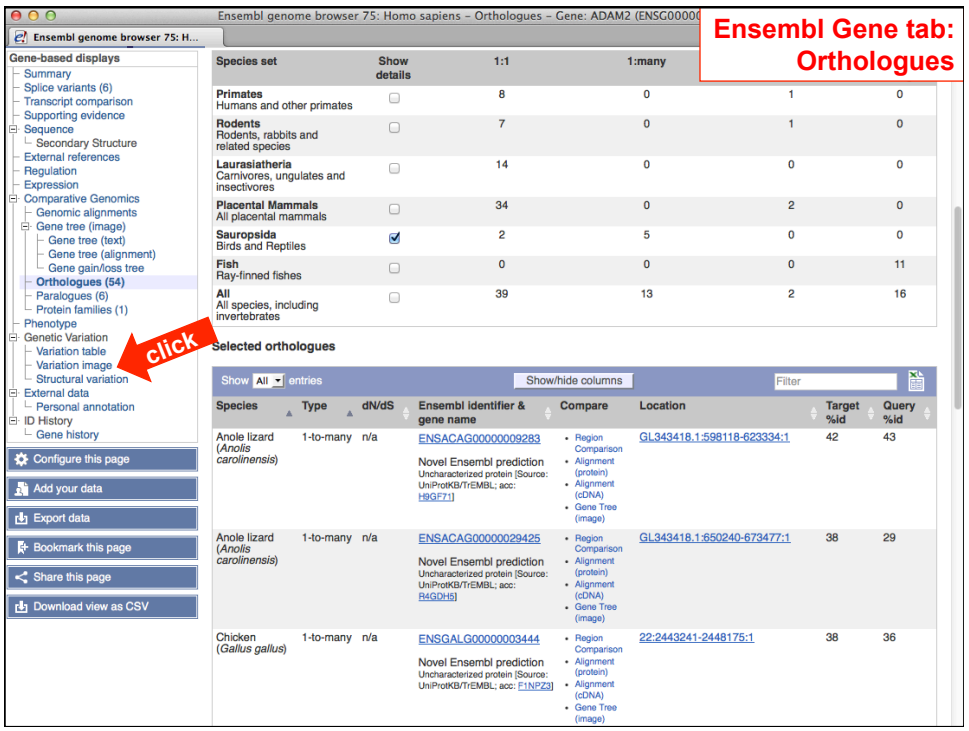

Ensembl Location tab:
Configure page
Region in detail
Configure Region Image
Configure Overview Image
Configure Chromosome Image
Personal Data
Default style:
BAM files
Gene models
Intron-spanning reads
Human BodyMap 2.0
Adipose
Adrenal
Blood
Brain
Breast
Colon
Heart
Kidney
Liver
Lung
Lymph
Merged (incl.Pooled)
Ovary
Pooled
Prostate
Skeletal muscle
Testes

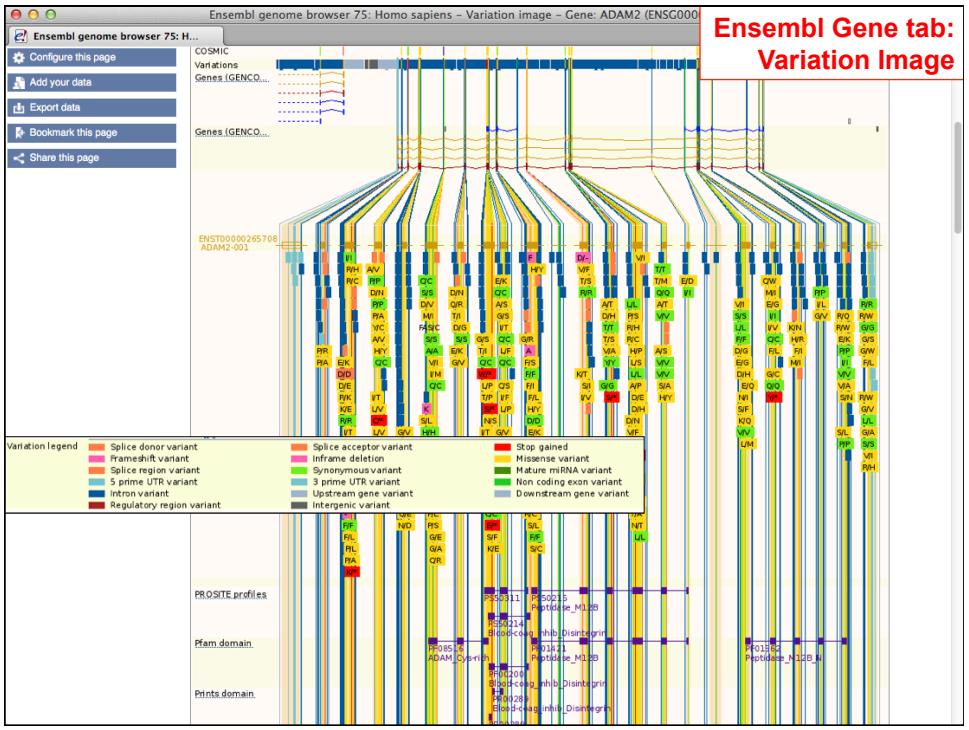

Ensembl Location tab:
RNASeq alignments
75: Homo sapiens – Region in detail – Chromosome 8: 39,588,397–39,638,396
Chromosome bands
37 way GERP ele...
Human cDNAs (R...
Testes RNAseq all
Skeletal muscle R...
Brain RNAseq alig...
Testes intron-spa...
Skeletal muscle i...
Brain intron-span...
Contigs
Genes (GENCODE...
CCDS set
Brain intron-span...
Skeletal muscle i...
Testes intron-spa...
Testes RNASeq
Human cDNAs (R...
Constrained elements for 37 eutherian mammals EPO_LOW_COVERAGE
5257 features from 'Testes RNAseq alignments' omitted
< AC136365.5
< AC136365.1-201
miRNA
< ADAM2-006
processed transcript
< ADAM2-003
protein coding
< ADAM2-004
protein coding
< ADAM2-001
protein coding
< ADAM2-002
protein coding
< CCDS64882.1
CCDS set
< CCDS34884.1
CCDS set
< CCDS64883.1
CCDS set


Ensembl
Gene tab
http://www.ensembl.org
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research


Ensembl Location tab:
Region in detail
Ensembl genome browser 75: Homo sapiens – Region in detail – Chromosome 8: 39
Ensembl genome browser 75: H...
NCBI
Vega
Configure this page
Add your data
Export data
Bookmark this page
Share this page
Chromosome bands
Contigs
Genes (GENCODE 19)
Gene Legend
Merged Ensembl/Havana
Processed transcript
Pseudogene
RNA gene
Location: 8:39613346-39613446
Gene:
Go
Chromosome bands
37 way GERP ele...
Sequence
Contigs
Sequence
Genes (GENCODE...
Constrained elements for 37 eutherian mammals EPO_LOW_COVERAGE
HGNC Symbol: ADAM2-001
Exon
Transcript
Protein
Gene
Location
Gene type
Transcript type
Strand
Base pairs
Amino acids
Analysis
Prediction method
click
CCDS set
Human cDNAs (R...
%GC
Blocks show the locations of G/C base pairs.
merged Ensembl/Havana


Ensembl Gene tab:
Gene summary
Ensembl genome browser 75: Homo sapiens – Summary – Gene: ADAM2 (ENSG00000
Human (GRCh37)
Location: 8:39,613,346-39,613,446
Gene: ADAM2
Transcript: ADAM2-001
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 (alignment)
Gene gain/loss tree
Orthologues (54)
Paralogues (6)
Protein families (1)
Phenotype
Genetic Variation
Variation table
Variation image
Structural variation
External data
Personal annotation
ID History
Gene history
click
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
Show/hide columns
Filter
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 ENSP00000429352 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
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: OTTHUMG00000164041 (version 3)
Go to Region in Detail for more tracks and navigation options (e.g. zooming)


Ensembl Gene tab:
Orthologues
Ensembl genome browser 75: Homo sapiens – Orthologues – Gene: ADAM2
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 (alignment)
Gene gain/loss tree
Orthologues (54)
Paralogues (6)
Protein families (1)
Phenotype
Genetic Variation
Variation table
Variation image
Structural variation
External data
Personal annotation
ID History
Gene history
click
Configure this page
Add your data
Export data
Bookmark this page
Share this page
Download view as CSV
Species set
Show details
1:1
1:many
Primates
Humans and other primates
Rodents
Rodents, rabbits and related species
Laurasiatheria
Carnivores, ungulates and insectivores
Placental Mammals
All placental mammals
Sauropsida
Birds and Reptiles
Fish
Ray-finned fishes
All
All species, including invertebrates
Selected orthologues
Show All entries
Show/hide columns
Filter
Species
Type
dN/dS
Ensembl identifier & gene name
Compare
Location
Target %id
Query %id
Anole lizard (Anolis carolinensis)
1-to-many
n/a
ENSACAG00000009283
Novel Ensembl prediction
GL343418.1:598118-623334:1
42
43
Anole lizard (Anolis carolinensis)
1-to-many
n/a
ENSACAG00000029425
Novel Ensembl prediction
GL343418.1:650240-673477:1
38
29
Chicken (Gallus gallus)
1-to-many
n/a
ENSGALG00000003444
Novel Ensembl prediction
22:2443241-2448175:1
38
36
Region Comparison
Alignment (protein)
Alignment (cDNA)
Gene Tree (image)

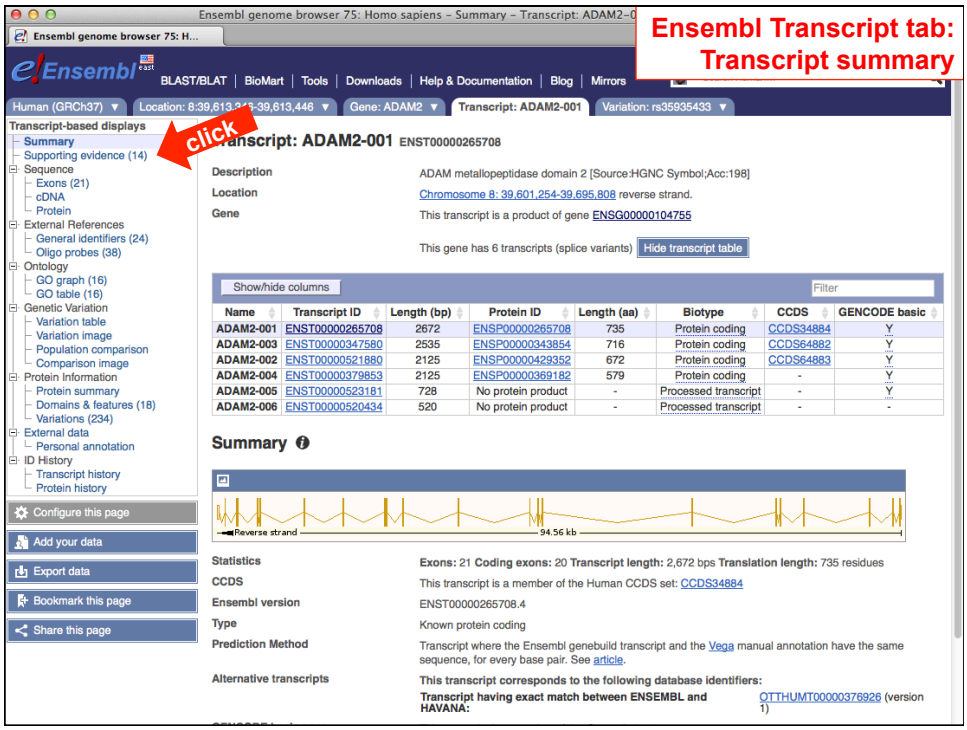

Ensembl Gene tab:
Variation Image
Ensembl genome browser 75: Homo sapiens – Variation image – Gene: ADAM2
Configure this page
Add your data
Export data
Bookmark this page
Share this page
COSMIC Variations
Genes (GENCO...
Variation legend
Splice donor variant
Frameshift variant
Splice region variant
5 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

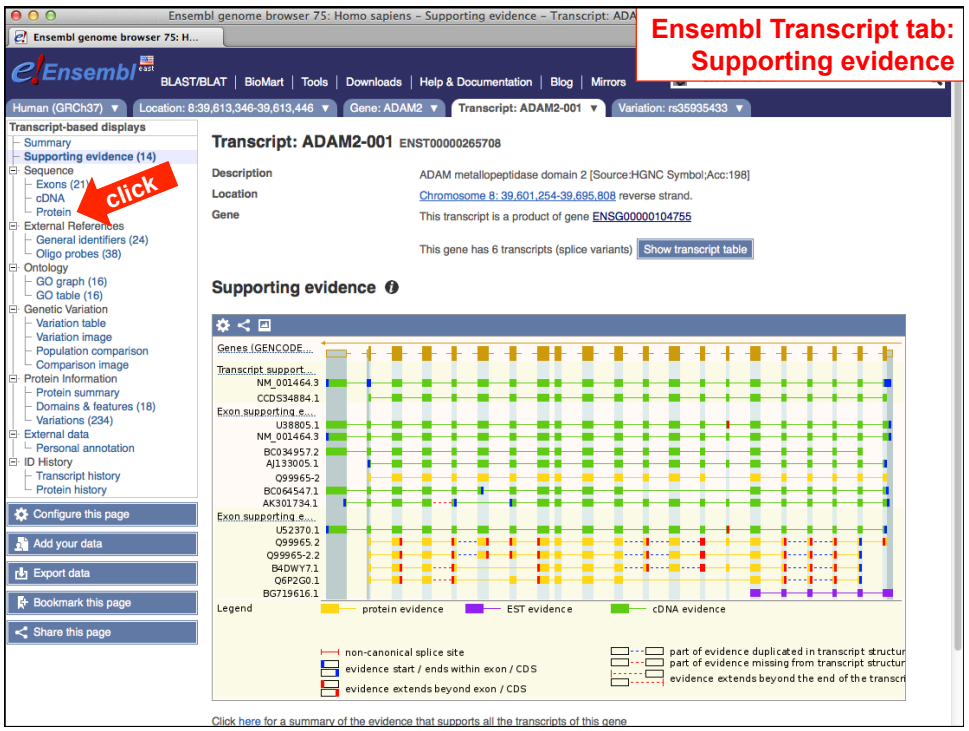
Ensembl

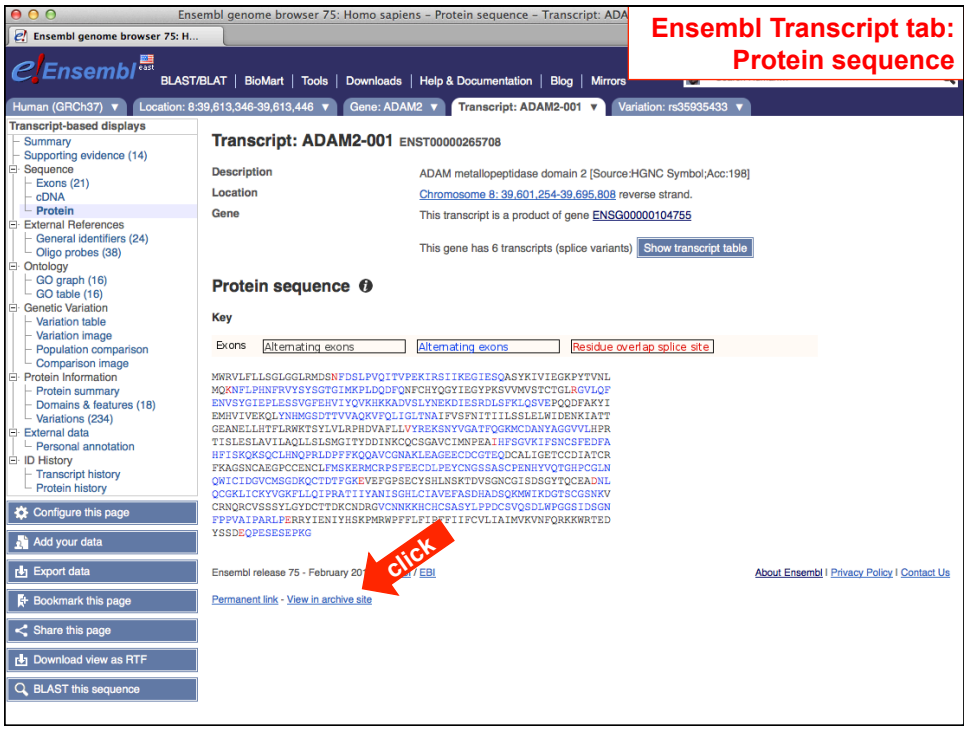
Transcript tab

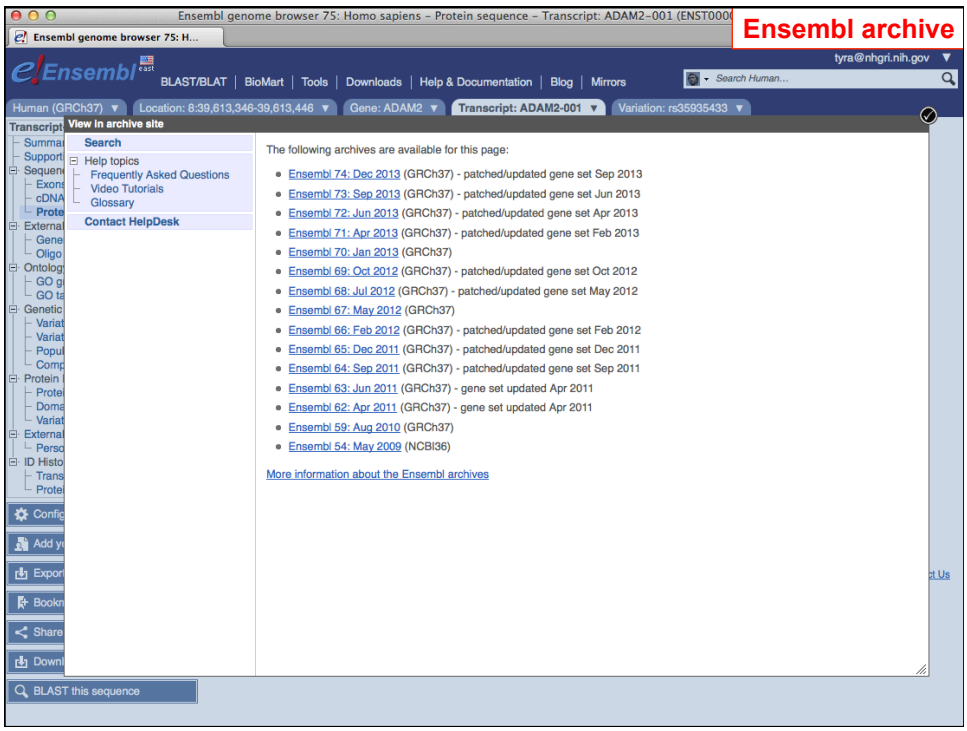
http://www.ensembl.org

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

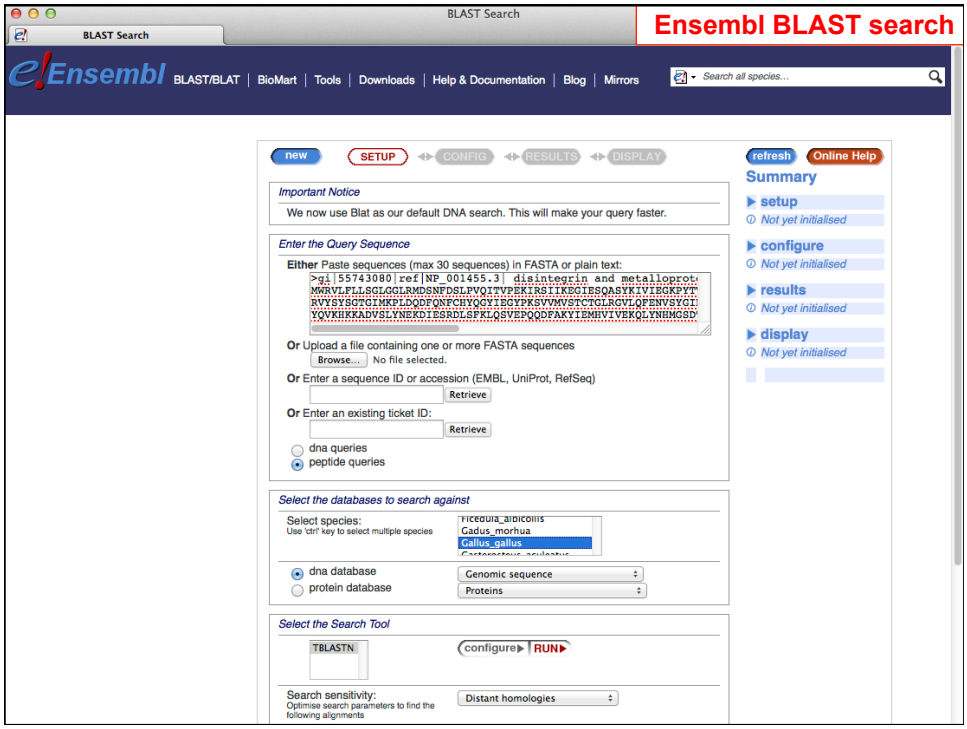


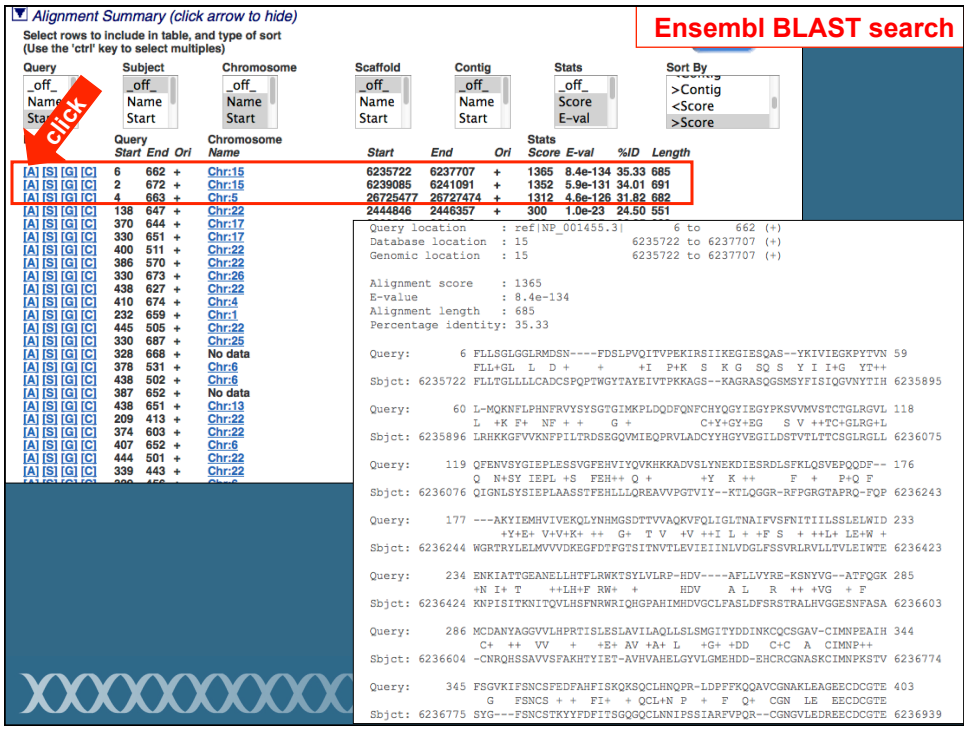

Ensembl Transcript tab:
Supporting evidence
Ensembl genome browser 75: Homo sapiens – Supporting evidence – Transcript: ADA
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Human (GRCh37) | Location: 8:39,613,346-39,613,446 | Gene: ADAM2 | Transcript: ADAM2-001 | Variation: rs35935433
Transcript-based displays
Summary
Supporting evidence (14)
Sequence
Exons (21)
cDNA
Protein
External References
General identifiers (24)
Oligo probes (38)
Ontology
GO graph (16)
GO table (16)
Genetic Variation
Variation table
Variation image
Population comparison
Comparison image
Protein Information
Protein summary
Domains & features (18)
Variations (234)
External data
Personal annotation
ID History
Transcript history
Protein history
click
Configure this page
Add your data
Export data
Bookmark this page
Share this page
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)
Show transcript table
Supporting evidence
Genes (GENCODE...
Transcript support...
NM_001464.3
CCDS34884.1
Exon supporting e...
U38805.1
NM_001464.3
BC034957.2
AJ133005.1
Q99965-2
BC064547.1
AK301734.1
Exon supporting e...
U52370.1
Q99965.2
Q99965-2.2
B4DWY7.1
Q6P2G0.1
BG719616.1
Legend
protein evidence
EST evidence
cDNA evidence
non-canonical splice site
evidence start / ends within exon / CDS
evidence extends beyond exon / CDS
part of evidence duplicated in transcript structur
part of evidence missing from transcript structur
evidence extends beyond the end of the transcri
Click here for a summary of the evidence that supports all the transcripts of this gene


Ensembl Transcript tab:
Protein sequence
Ensembl genome browser 75: Homo sapiens – Protein sequence – Transcript: ADA
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Human (GRCh37) | Location: 8:39,613,346-39,613,446 | Gene: ADAM2 | Transcript: ADAM2-001 | Variation: rs35935433
Transcript-based displays
Summary
Supporting evidence (14)
Sequence
Exons (21)
cDNA
Protein
External References
General identifiers (24)
Oligo probes (38)
Ontology
GO graph (16)
GO table (16)
Genetic Variation
Variation table
Variation image
Population comparison
Comparison image
Protein Information
Protein summary
Domains & features (18)
Variations (234)
External data
Personal annotation
ID History
Transcript history
Protein history
Configure this page
Add your data
Export data
Bookmark this page
Share this page
Download view as RTF
BLAST this sequence
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)
Show transcript table
Protein sequence
Key
Exons
Alternating exons
Alternating exons
Residue overlap splice site
MWRVLFLLSGLGGLRMDSNFDSLPVQITVPEKIRSIIKEGIESQASYKIVIEGKPYTVNL
MQKNFLPHNFRVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQF
ENVSYGIEPLESSVGFEHVIYQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDFAKYI
EMHVIVEKQLYNHMGSDTTVVAQKVFQLIGLTNAIFVSFNITIILSSLELWIDENKIATT
GEANELLHTFLRWKTSYLVLRPHDVAFLLVYREKSNYVGATFQGKMCDANYAGGVVLHPR
TISLESLAVILAQLLSLSMGITYDDINKCQCSGAVCIMNPEAIHFSGVKIFSNCSFEDFA
HFISKQKSQCLHNQPRLDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR
FKAGSNCAEGPCCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLN
QWICIDGVCMSGDKQCTDTFGKEVEFGPSECYSHLNSKTDVSGNCGISDSGYTQCEADNL
QCGKLICKYVGKFLLQIPRATIIYANISGHLCIAVEFASDHADSQKMWIKDGTSCGSNKV
CRNQRCVSSSYLGYDCTPDKCNDRGVCNNKKHCHCSASYLPPDCSVQSDLWPGGSIDSGN
FPPVAIPARLPERRYIENIYHSKPMRWPFFLFIPFFIIPCVLIAIMVKVNFQRKKWRTED
YSSDEQPESESEPKG
Ensembl release 75 - February 201
EBI
About Ensembl | Privacy Policy | Contact Us
click
Permanent link - View in archive site


Ensembl archive
Ensembl genome browser 75: Homo sapiens – Protein sequence – Transcript: ADAM2-001
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
View in archive site
Search
Help topics
Frequently Asked Questions
Video Tutorials
Glossary
Contact HelpDesk
The following archives are available for this page:
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)
More information about the Ensembl archives
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


Ensembl BLAST search
BLAST Search
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Search all species...
new
SETUP
CONFIG
RESULTS
DISPLAY
refresh
Online Help
Summary
setup
Not yet initialised
configure
Not yet initialised
results
Not yet initialised
display
Not yet initialised
Important Notice
We now use Blat as our default DNA search. This will make your query faster.
Enter the Query Sequence
Either Paste sequences (max 30 sequences) in FASTA or plain text:
Or Upload a file containing one or more FASTA sequences
Browse...
No file selected.
Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)
Retrieve
Or Enter an existing ticket ID:
Retrieve
dna queries
peptide queries
Select the databases to search against
Select species:
Use 'ctrl' key to select multiple species
Gadus_morhua
Gallus_gallus
dna database
protein database
Genomic sequence
Proteins
Select the Search Tool
TBLASTN
configure
RUN
Search sensitivity:
Optimise search parameters to find the following alignments
Distant homologies

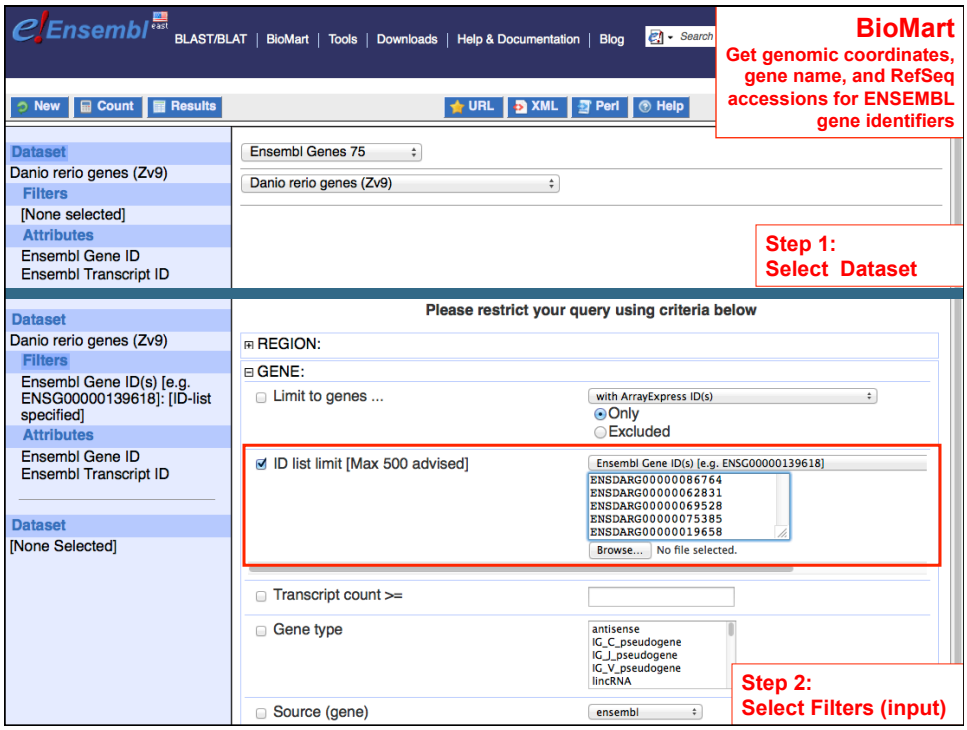

Ensembl BLAST search
Alignment Summary (click arrow to hide)
Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)
Query
Subject
Chromosome
Scaffold
Contig
Stats
Sort By
>Contig
<Score
>Score
click
Query Start End Ori
Chromosome Name
Start End Ori Score E-val %ID 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] 4 663 + 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
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 FLLSGLGGLRMDSN----FDSLPVQITVPEKIRSIIKEGIESQAS--YKIVIEGKPYTVN 59
Sbjct: 6235722 FLLTGLLLLCADCSPQPTWGYTAYEIVTPKKAGS--KAGRASQGSMSYFISIQGVNYTIH 6235895
Query: 60 L-MQKNFLPHNFRVYSYSGTGIMKPLDQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVL 118
Sbjct: 6235896 LRHKKGFVVKNFPILTRDSEGQVMIEQPRVLADCYYHGYVEGILDSTVTLTTCSGLRGLL 6236075
Query: 119 QFENVSYGIEPLESSVGFEHVIYQVKHKKADVSLYNEKDIESRDLSFKLQSVEPQQDF-- 176
Sbjct: 6236076 QIGNLSYSIEPLAASSTFEHLLLQREAVVPGTVIY--KTLQGGR-RFPGRGTAPRQ-FQP 6236243
Query: 177 ---AKYIEMHVIVEKQLYNHMGSDTTVVAQKVFQLIGLTNAIFVSFNITIILSSLELWID 233
Sbjct: 6236244 WGRTRYLELMVVVDKEGFDTFGTSITNVTLEVIEIINLVDGLFSSVRLRVLLTVLEIWTE 6236423
Query: 234 ENKIATTGEANELLHTFLRWKTSYLVLRP-HDV----AFLLVYRE-KSNYVG--ATFQGK 285
Sbjct: 6236424 KNPISITKNITQVLHSFNRWRIQHGPAHIMHDVGCLFASLDFSRSTRALHVGGESNFASA 6236603
Query: 286 MCDANYAGGVVLHPRTISLESLAVILAQLLSLSMGITYDDINKCQCSGAV-CIMNPEAIH 344
Sbjct: 6236604 -CNRQHSSAVVSFAKHTYIET-AVHVAHELGYVLGMEHDD-EHCRCGNASKCIMNPKSTV 6236774
Query: 345 FSGVKIFSNCSFEDFAHFISKQKSQCLHNQPR-LDPFFKQQAVCGNAKLEAGEECDCGTE 403
Sbjct: 6236775 SYG---FSNCSTKYYFDFITSGQGQCLNNIPSSIARFVPQR--CGNGVLEDREECDCGTE 6236939

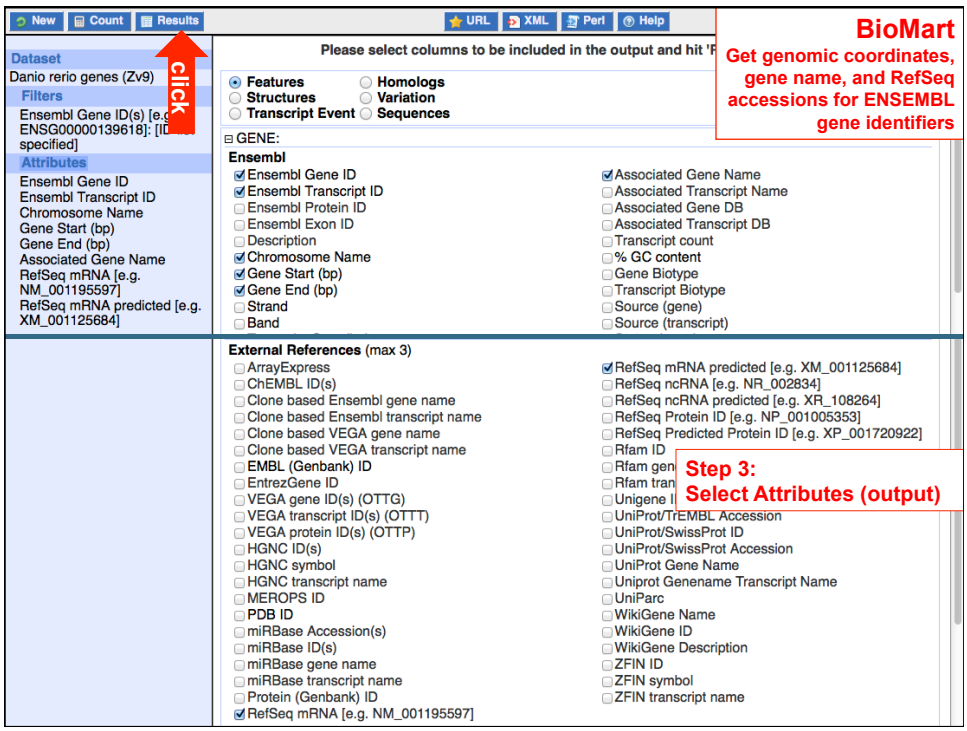
BioMart

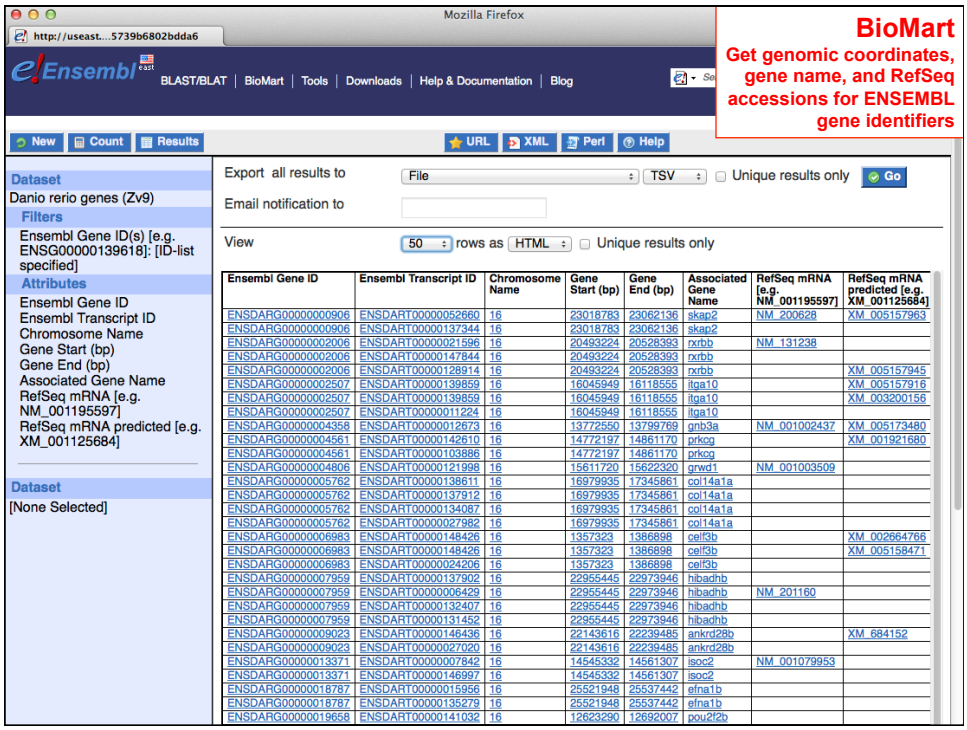
Cross-reference data from different sources

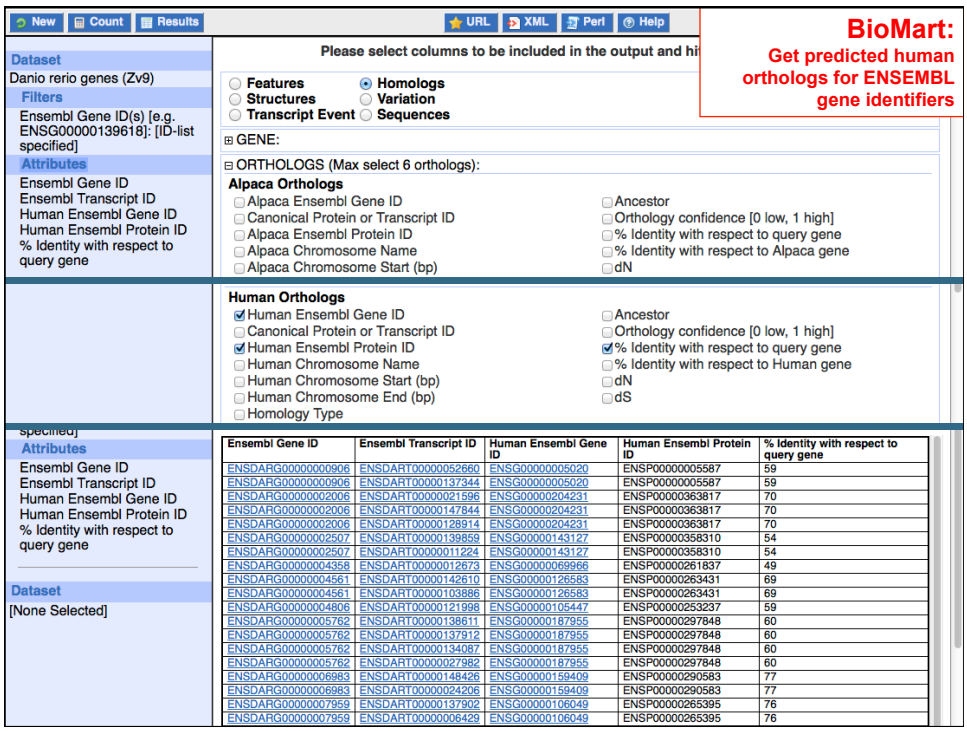
http://www.ensembl.org/biomart

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

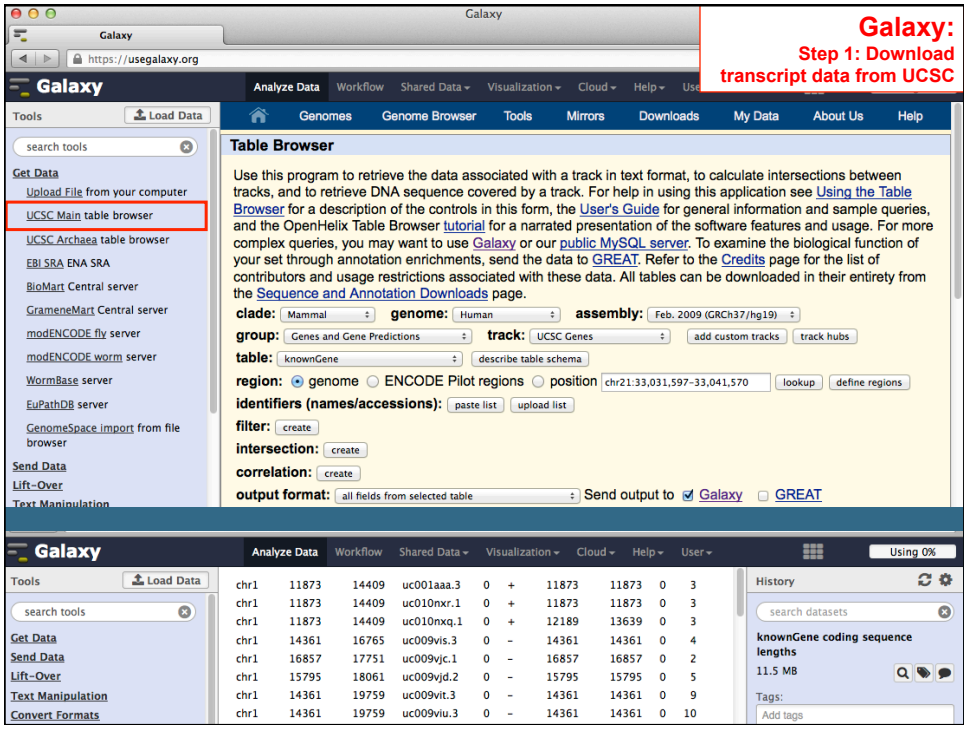


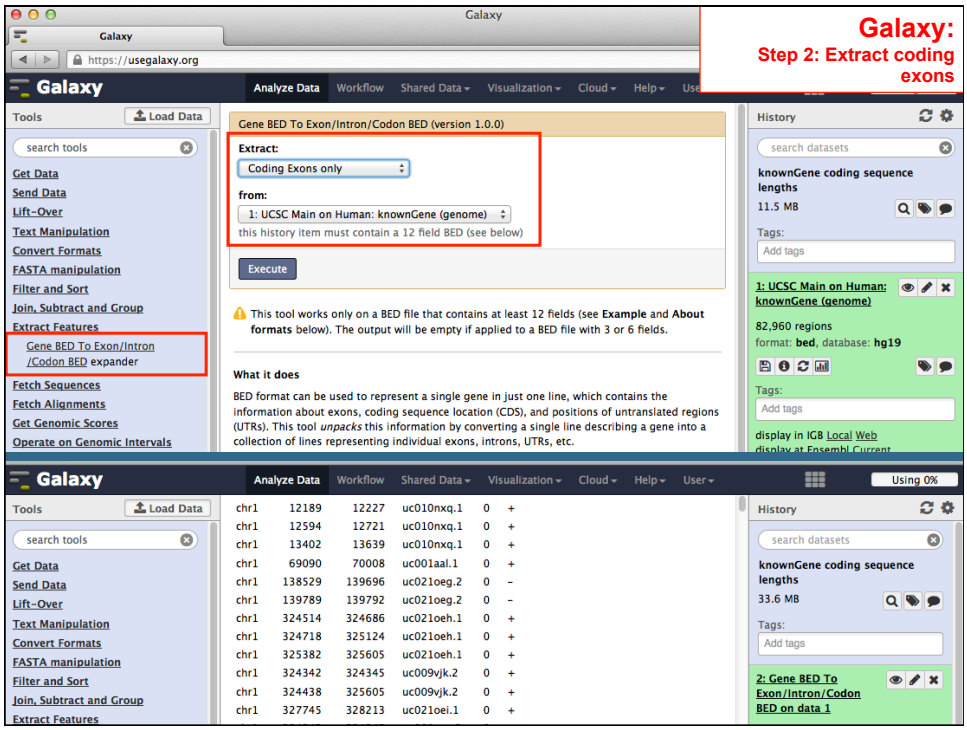

New
Count
Results
URL
XML
Perl
Help
Dataset
Danio rerio genes (Zv9)
Filters
Ensembl Gene ID(s) [e.g. ENSG00000139618]: [ID-list specified]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Chromosome Name
Gene Start (bp)
Gene End (bp)
Associated Gene Name
RefSeq mRNA [e.g. NM_001195597]
RefSeq mRNA predicted [e.g. XM_001125684]
click
Please select columns to be included in the output and hit 'F
Features
Structures
Transcript Event
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
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

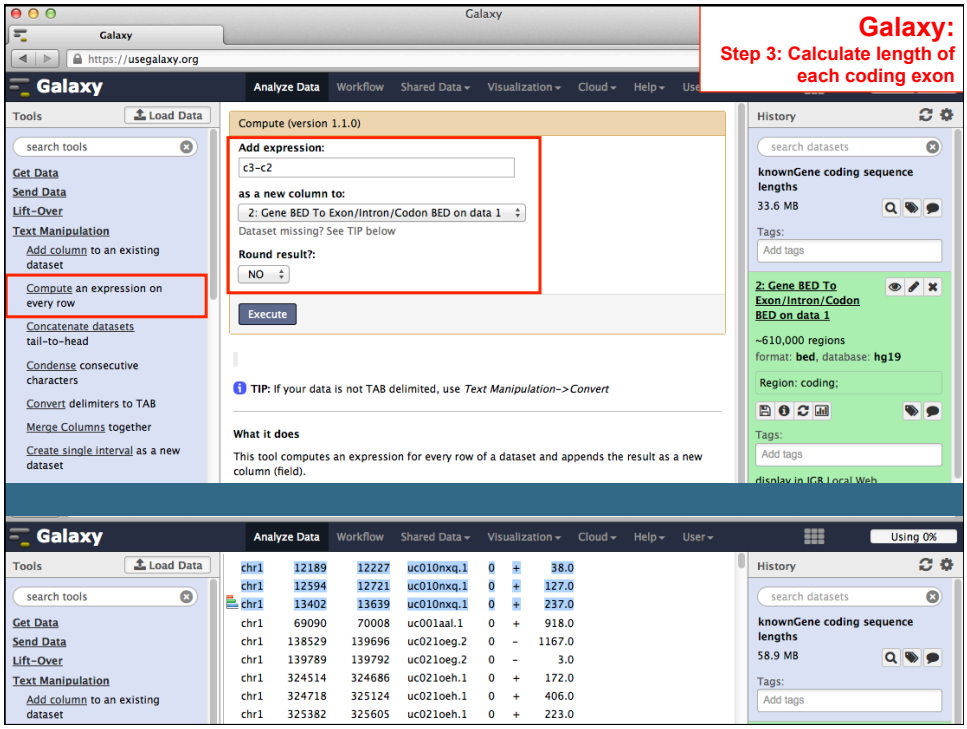

Mozilla Firefox
http://useast...5739b6802bdda6
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog
New
Count
Results
URL
XML
Perl
Help
Dataset
Danio rerio genes (Zv9)
Filters
Ensembl Gene ID(s) [e.g. ENSG00000139618]: [ID-list specified]
Attributes
Ensembl Gene ID
Ensembl Transcript 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]
Dataset
[None Selected]
Export all results to
File
TSV
Unique results only
Go
Email notification to
View
50
rows as
HTML
Unique results only
Ensembl Gene ID | Ensembl Transcript ID | 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 | ENSDART00000052660 | 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 | itga10 | | XM_005157916
ENSDARG00000002507 | ENSDART00000139859 | 16 | 16045949 | 16118555 | itga10 | | XM_003200156
ENSDARG00000002507 | ENSDART00000011224 | 16 | 16045949 | 16118555 | itga10
ENSDARG00000004358 | ENSDART00000012673 | 16 | 13772550 | 13799769 | gnb3a | NM_001002437 | XM_005173480
ENSDARG00000004561 | ENSDART00000142610 | 16 | 14772197 | 14861170 | prkcg | | XM_001921680
ENSDARG00000004561 | ENSDART00000103886 | 16 | 14772197 | 14861170 | prkcg
ENSDARG00000004806 | ENSDART00000121998 | 16 | 15611720 | 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 | celf3b | | XM_002664766
ENSDARG00000006983 | ENSDART00000148426 | 16 | 1357323 | 1386898 | celf3b | | XM_005158471
ENSDARG00000006983 | ENSDART00000024206 | 16 | 1357323 | 1386898 | celf3b
ENSDARG00000007959 | ENSDART00000137902 | 16 | 22955445 | 22973946 | hibadhb
ENSDARG00000007959 | ENSDART00000006429 | 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 | 25521948 | 25537442 | efna1b
ENSDARG00000018787 | ENSDART00000135279 | 16 | 25521948 | 25537442 | efna1b
ENSDARG00000019658 | ENSDART00000141032 | 16 | 12623290 | 12692007 | pou2f2b
BioMart
Get genomic coordinates, gene name, and RefSeq accessions for ENSEMBL gene identifiers

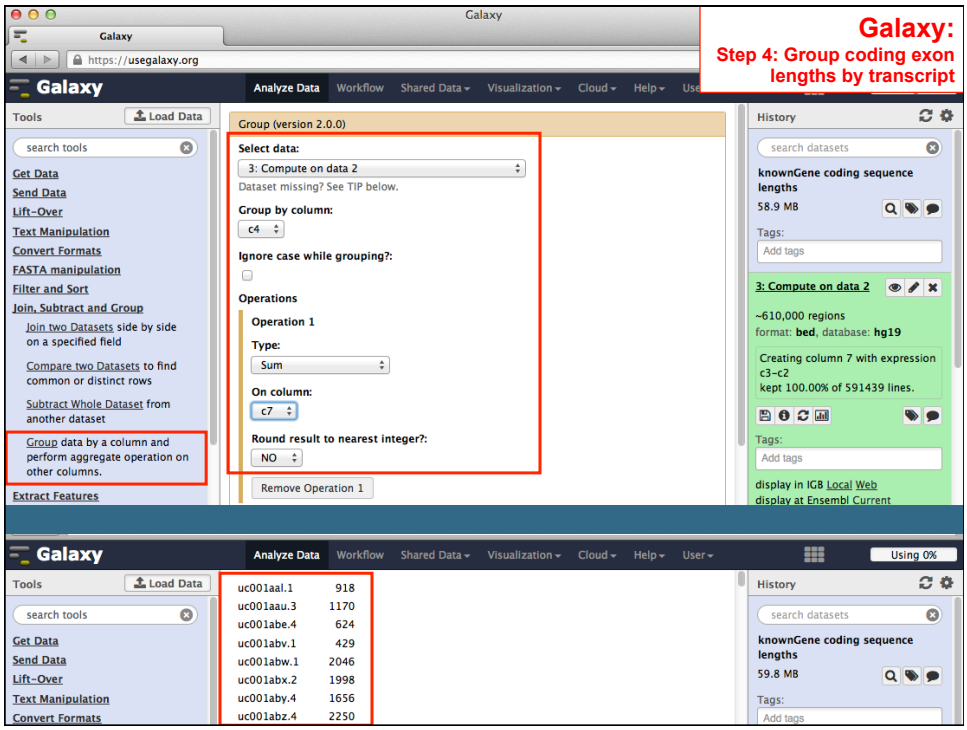

New
Count
Results
URL
XML
Perl
Help
BioMart:
Get predicted human orthologs for ENSEMBL gene identifiers
Dataset
Danio rerio genes (Zv9)
Filters
Ensembl Gene ID(s) [e.g. ENSG00000139618]: [ID-list specified]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Human Ensembl Gene ID
Human Ensembl Protein ID
% Identity with respect to query gene
Please select columns to be included in the output and hi
Features
Homologs
Structures
Variation
Transcript Event
Sequences
GENE:
ORTHOLOGS (Max select 6 orthologs):
Alpaca Orthologs
Alpaca Ensembl Gene ID
Alpaca Ensembl Protein ID
Alpaca Chromosome Name
Alpaca Chromosome Start (bp)
Ancestor
Orthology confidence [0 low, 1 high]
% Identity with respect to query gene
% Identity with respect to Alpaca gene
dN
Human Orthologs
Human Ensembl Gene ID
Human Ensembl Protein ID
Human Chromosome Name
Human Chromosome Start (bp)
Human Chromosome End (bp)
Homology Type
Ancestor
Orthology confidence [0 low, 1 high]
% Identity with respect to query gene
% Identity with respect to Human gene
dN
dS
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Human Ensembl Gene ID
Human Ensembl Protein ID
% Identity with respect to query gene
Dataset
[None Selected]
Ensembl Gene ID
Ensembl Transcript ID
Human Ensembl Gene ID
Human Ensembl Protein ID
% Identity with respect to query gene
ENSDARG00000000906 ENSDART00000052660 ENSG00000005020 ENSP00000005587 59
ENSDARG00000000906 ENSDART00000137344 ENSG00000005020 ENSP00000005587 59
ENSDARG00000002006 ENSDART00000021596 ENSG00000204231 ENSP00000363817 70
ENSDARG00000002006 ENSDART00000147844 ENSG00000204231 ENSP00000363817 70
ENSDARG00000002006 ENSDART00000128914 ENSG00000204231 ENSP00000363817 70
ENSDARG00000002507 ENSDART00000139859 ENSG00000143127 ENSP00000358310 54
ENSDARG00000002507 ENSDART00000011224 ENSG00000143127 ENSP00000358310 54
ENSDARG00000004358 ENSDART00000012673 ENSG00000069966 ENSP00000261837 49
ENSDARG00000004561 ENSDART00000142610 ENSG00000126583 ENSP00000263431 69
ENSDARG00000004561 ENSDART00000103886 ENSG00000126583 ENSP00000263431 69
ENSDARG00000004806 ENSDART00000121998 ENSG00000105447 ENSP00000253237 59
ENSDARG00000005762 ENSDART00000138611 ENSG00000187955 ENSP00000297848 60
ENSDARG00000005762 ENSDART00000137912 ENSG00000187955 ENSP00000297848 60
ENSDARG00000005762 ENSDART00000134087 ENSG00000187955 ENSP00000297848 60
ENSDARG00000005762 ENSDART00000027982 ENSG00000187955 ENSP00000297848 60
ENSDARG00000006983 ENSDART00000148426 ENSG00000159409 ENSP00000290583 77
ENSDARG00000006983 ENSDART00000024206 ENSG00000159409 ENSP00000290583 77
ENSDARG00000007959 ENSDART00000137902 ENSG00000106049 ENSP00000265395 76
ENSDARG00000007959 ENSDART00000006429 ENSG00000106049 ENSP00000265395 76

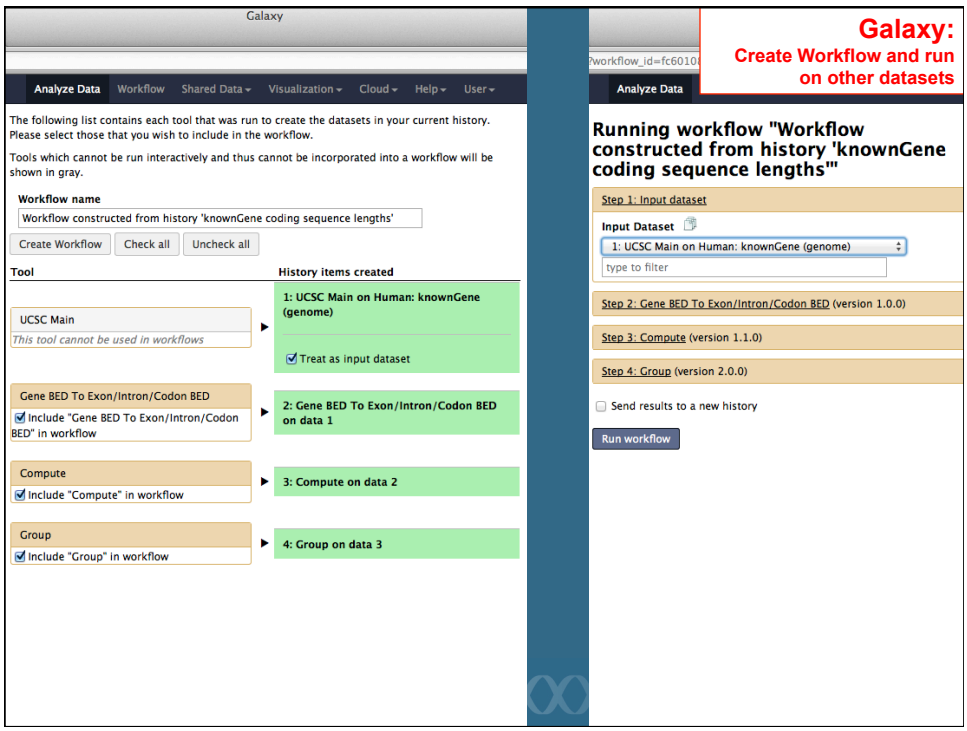

Galaxy
knownGene coding sequence lengths
https://usegalaxy.org/
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research


Galaxy:
Step 1: Download transcript data from UCSC
Galaxy
https://usegalaxy.org
Analyze Data
Workflow
Shared Data
Visualization
Cloud
Help
Tools
Load Data
search tools
Get Data
Upload File from your computer
UCSC Main table browser
UCSC Archaea 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
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
About Us
Help
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
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
Using 0%
History
search datasets
knownGene coding sequence lengths
11.5 MB
Tags:
Add tags


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)
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.
Gene BED To Exon/Intron /Codon BED expander
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
FASTA manipulation
Filter and Sort
Join, Subtract and Group
Extract Features
1: UCSC Main on Human: knownGene (genome)
82,960 regions
format: bed, database: hg19
display in IGB Local Web
2: Gene BED To Exon/Intron/Codon BED on data 1
33.6 MB


Galaxy:
Step 3: Calculate length of each coding exon
Galaxy
https://usegalaxy.org
Analyze Data
Workflow
Shared Data
Visualization
Cloud
Help
Tools
Load Data
search tools
Get Data
Send Data
Lift-Over
Text Manipulation
Add column to an existing dataset
Compute an expression on every row
Concatenate datasets tail-to-head
Condense consecutive characters
Convert delimiters to TAB
Merge Columns together
Create single interval as a new dataset
Compute (version 1.1.0)
Add expression:
c3-c2
as a new column to:
2: Gene BED To Exon/Intron/Codon BED on data 1
Dataset missing? See TIP below
Round result?:
NO
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).
History
search datasets
knownGene coding sequence lengths
33.6 MB
Tags:
Add tags
2: Gene BED To Exon/Intron/Codon BED on data 1
~610,000 regions
format: bed, database: hg19
Region: coding;
Using 0%
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
58.9 MB


Galaxy:
Step 4: Group coding exon lengths by transcript
Galaxy
https://usegalaxy.org
Analyze Data
Workflow
Shared Data
Visualization
Cloud
Help
Tools
Load Data
search tools
Get Data
Send Data
Lift-Over
Text Manipulation
Convert Formats
FASTA manipulation
Filter and Sort
Join, Subtract and Group
Join two Datasets side by side on a specified field
Compare two Datasets to find common or distinct rows
Subtract Whole Dataset from another dataset
Group data by a column and perform aggregate operation on other columns.
Extract Features
Group (version 2.0.0)
Select data:
3: Compute on data 2
Dataset missing? See TIP below.
Group by column:
c4
Ignore case while grouping?:
Operations
Operation 1
Type:
Sum
On column:
c7
Round result to nearest integer?:
NO
Remove Operation 1
History
search datasets
knownGene coding sequence lengths
58.9 MB
Tags:
Add tags
3: Compute on data 2
~610,000 regions
format: bed, database: hg19
Creating column 7 with expression c3-c2
kept 100.00% of 591439 lines.
display in IGB Local Web
display at Ensembl Current
Using 0%
uc001aal.1 918
uc001aau.3 1170
uc001abe.4 624
uc001abv.1 429
uc001abw.1 2046
uc001abx.2 1998
uc001aby.4 1656
uc001abz.4 2250
59.8 MB



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

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

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

The UCSC Genome Browser — *UNIT 1.4*

Donna Karolchik,[1] Angie S. Hinrichs,[1] and W. James Kent[1]

[1]Center for Biomolecular Science and Engineering, University of Calif
Santa Cruz, California

ABSTRACT

The University of California Santa Cruz (UCSC) Genome Brows
based tool for quickly displaying a requested portion of a genor
companied by a series of aligned annotation "tracks." The annotati
UCSC Genome Bioinformatics Group and external collaborators i
tions, mRNA and expressed sequence tag alignments, simple nucleo
expression and regulatory data, phenotype and variation data, and pa
species comparative genomics data. All information relevant to a r
one window, facilitating biological analysis and interpretation. The
derlying the Genome Browser tracks can be viewed, downloaded, ar
another Web-based application, the UCSC Table Browser. Users

Using Galaxy to Perform Large-Scale Interactive Data Analyses — *UNIT 10.5*

Jennifer Hillman-Jackson,[1] Dave Clements,[2] Daniel Blankenberg,[1] James Taylor,[2] Anton Nekrutenko,[1] and Galaxy Team[1,2]

[1]Penn State University, University Park, Pennsylvania
[2]Emory University, Atlanta, Georgia

ABSTRACT

Innovations in biomedical research technologies continue to provide experimental biol-
ogists 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?
up and use is riddled with complexities outside of
e authors believe that Galaxy provides a powerful
on and analysis in an intuitive Web application,
nformatics tools previously only available to com-
based environments. We will demonstrate through
cols how Galaxy specifically brings together (1)
sources, for example, UCSC's Eukaryote and
om tools (wrapped Unix functions, format stan-
tions), and 3rd-party analysis tools. *Curr. Protoc.*
by John Wiley & Sons, Inc.

mics • genomic alignments •

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

Xosé M. Fernández-Suárez[1] and Michael K. Schuster[1]

[1]EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

ABSTRACT

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 genomic

Protoc. Bioinform. 30:1.15.1-1.15.48. © 2010 by John Wiley & Sons, Inc.

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

Access from NIH at
http://onlinelibrary.wiley.com/book/10.1002/0471250953