NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research

Current Topics in Genome Analysis 2016

Week 3: 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 2016

Tyra Wolfsberg, Ph.D.

No Relevant Financial Relationships with Commercial Interests

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

- UCSC Genome Browser
  http://genome.ucsc.edu
- Ensembl
  http://www.ensembl.org
- BioMart
  http://www.ensembl.org/biomart
- Integrative Genomics Viewer (IGV)
  http://www.broadinstitute.org/igv
- JBrowse
  http://jbrowse.org

Genome Browsers

- Exome Aggregation Consortium
  http://exac.broadinstitute.org

Variant Browser

- Galaxy
  https://usegalaxy.org

Web-based analysis

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

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

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Genome Sequence Assemblies

- Complex algorithms needed to incorporate all sequence data
- Assemblies updated periodically as new sequence becomes available
  - Mouse, human, and zebrafish 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

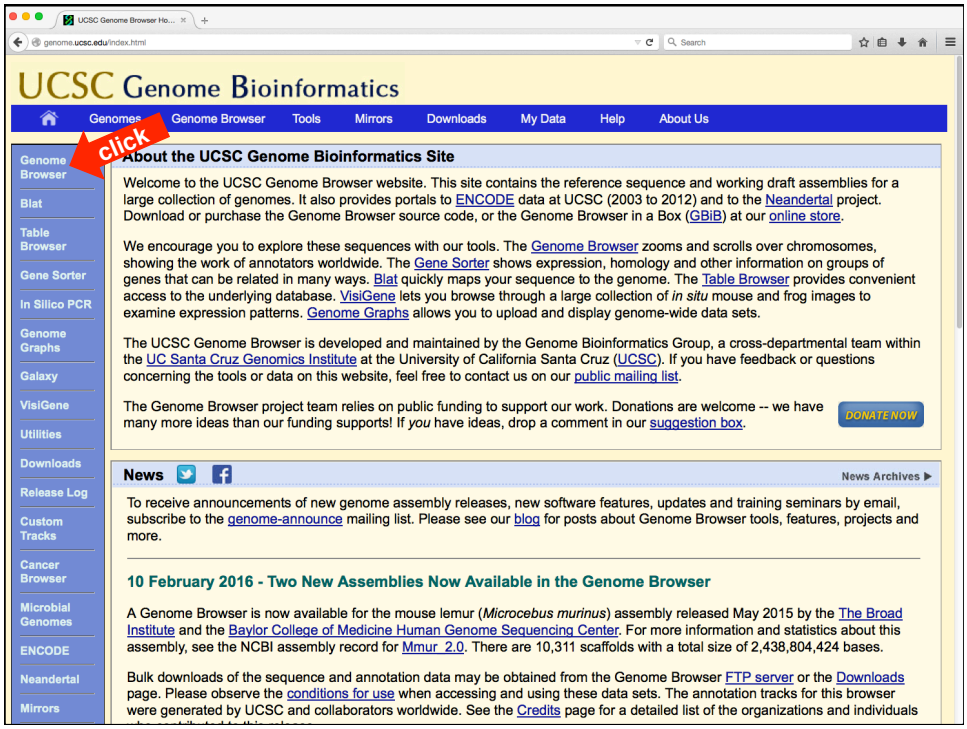
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

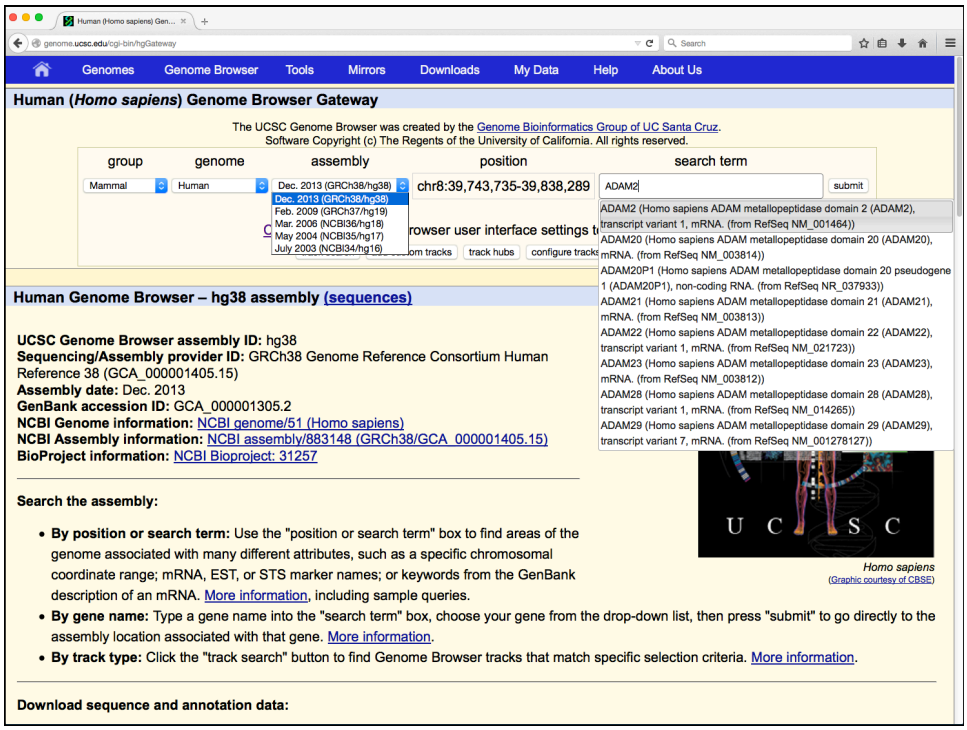
UCSC

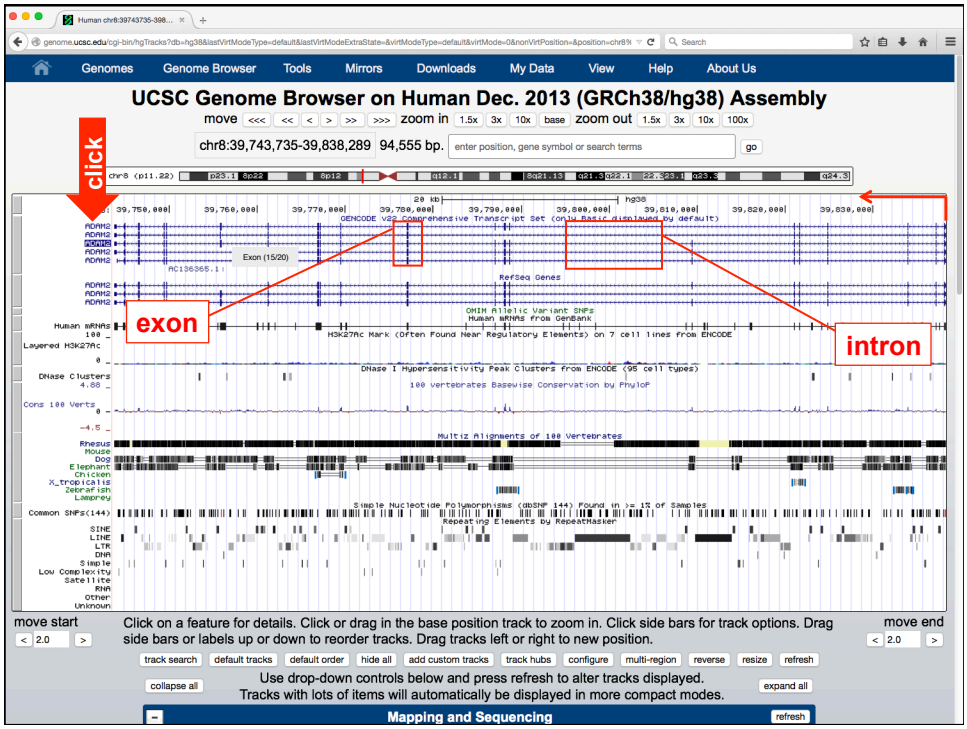
View a region in the genome by querying with a gene symbol

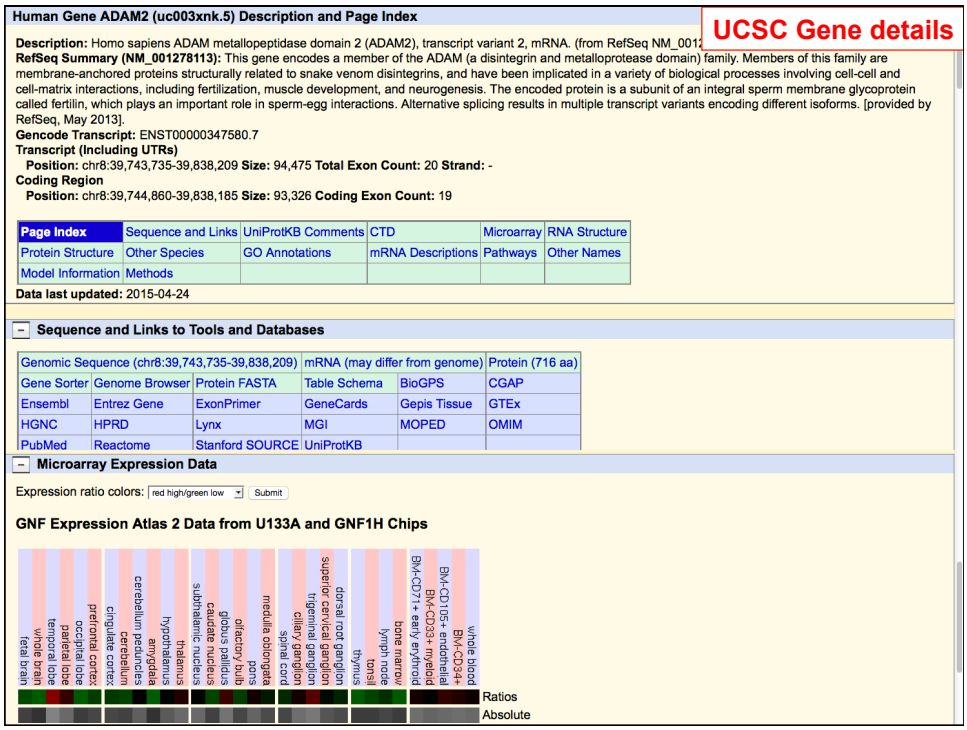
http://genome.ucsc.edu

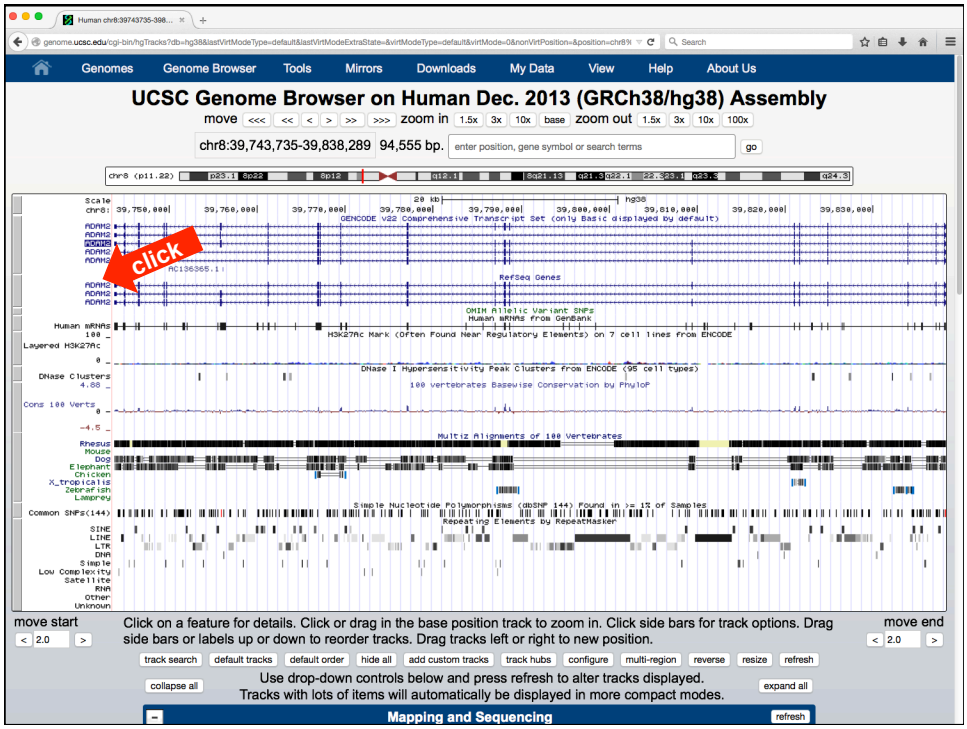
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

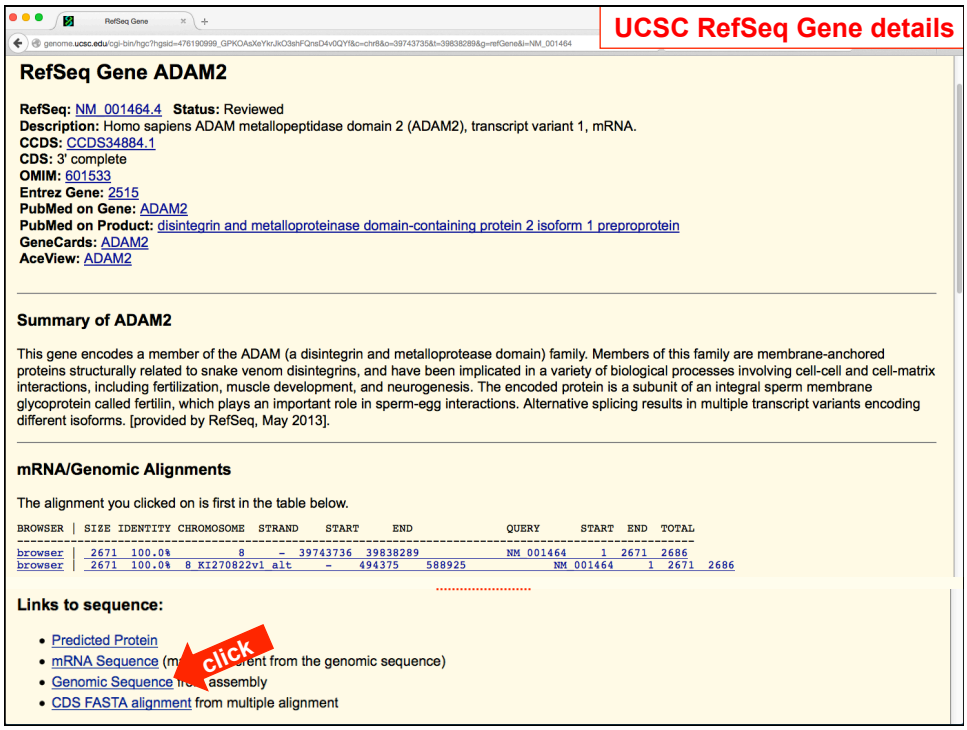

UCSC Genome Bioinformatics
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
Help
About Us
click
About the UCSC Genome Bioinformatics Site
Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to ENCODE data at UCSC (2003 to 2012) and to the Neandertal project. Download or purchase the Genome Browser source code, or the Genome Browser in a Box (GBiB) at our online store.
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 UC Santa Cruz Genomics Institute 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.
The Genome Browser project team relies on public funding to support our work. Donations are welcome -- we have many more ideas than our funding supports! If *you* have ideas, drop a comment in our suggestion box.
DONATE NOW
Genome Browser
Blat
Table Browser
Gene Sorter
In Silico PCR
Genome Graphs
Galaxy
VisiGene
Utilities
Downloads
Release Log
Custom Tracks
Cancer Browser
Microbial Genomes
ENCODE
Neandertal
Mirrors
News
News Archives
To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the genome-announce mailing list. Please see our blog for posts about Genome Browser tools, features, projects and more.
10 February 2016 - Two New Assemblies Now Available in the Genome Browser
A Genome Browser is now available for the mouse lemur (*Microcebus murinus*) assembly released May 2015 by the The Broad Institute and the Baylor College of Medicine Human Genome Sequencing Center. For more information and statistics about this assembly, see the NCBI assembly record for Mmur_2.0. There are 10,311 scaffolds with a total size of 2,438,804,424 bases.
Bulk downloads of the sequence and annotation data may be obtained from the Genome Browser FTP server or the Downloads page. Please observe the conditions for use when accessing and using these data sets. The annotation tracks for this browser were generated by UCSC and collaborators worldwide. See the Credits page for a detailed list of the organizations and individuals

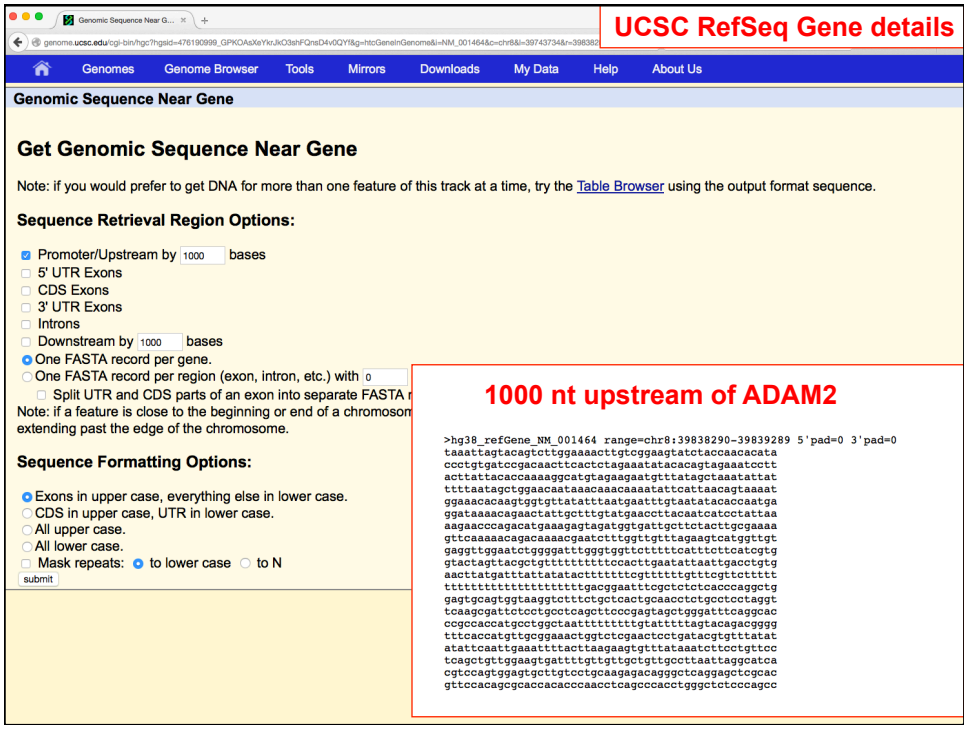

Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
Help
About Us
Human (*Homo sapiens*) Genome Browser Gateway
The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.
group
genome
assembly
position
search term
Mammal
Human
Dec. 2013 (GRCh38/hg38)
Feb. 2009 (GRCh37/hg19)
Mar. 2006 (NCBI36/hg18)
May 2004 (NCBI35/hg17)
July 2003 (NCBI34/hg16)
chr8:39,743,735-39,838,289
ADAM2
submit
track hubs
ADAM2 (Homo sapiens ADAM metallopeptidase domain 2 (ADAM2), transcript variant 1, mRNA. (from RefSeq NM_001464))
ADAM20 (Homo sapiens ADAM metallopeptidase domain 20 (ADAM20), mRNA. (from RefSeq NM_003814))
ADAM20P1 (Homo sapiens ADAM metallopeptidase domain 20 pseudogene 1 (ADAM20P1), non-coding RNA. (from RefSeq NR_037933))
ADAM21 (Homo sapiens ADAM metallopeptidase domain 21 (ADAM21), mRNA. (from RefSeq NM_003813))
ADAM22 (Homo sapiens ADAM metallopeptidase domain 22 (ADAM22), transcript variant 1, mRNA. (from RefSeq NM_021723))
ADAM23 (Homo sapiens ADAM metallopeptidase domain 23 (ADAM23), mRNA. (from RefSeq NM_003812))
ADAM28 (Homo sapiens ADAM metallopeptidase domain 28 (ADAM28), transcript variant 1, mRNA. (from RefSeq NM_014265))
ADAM29 (Homo sapiens ADAM metallopeptidase domain 29 (ADAM29), transcript variant 7, mRNA. (from RefSeq NM_001278127))
Human Genome Browser – hg38 assembly (sequences)
UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: GRCh38 Genome Reference Consortium Human Reference 38 (GCA_000001405.15)
Assembly date: Dec. 2013
GenBank accession ID: GCA_000001305.2
NCBI Genome information: NCBI genome/51 (Homo sapiens)
NCBI Assembly information: NCBI assembly/883148 (GRCh38/GCA_000001405.15)
BioProject information: NCBI Bioproject: 31257
U C S C
Homo sapiens
(Graphic courtesy of CBSE)
Search the assembly:
- **By position or search term:** Use the "position or search term" box to find areas of the genome associated with many different attributes, such as a specific chromosomal coordinate range; mRNA, EST, or STS marker names; or keywords from the GenBank description of an mRNA. More information, including sample queries.
- **By gene name:** Type a gene name into the "search term" box, choose your gene from the drop-down list, then press "submit" to go directly to the assembly location associated with that gene. More information.
- **By track type:** Click the "track search" button to find Genome Browser tracks that match specific selection criteria. More information.
Download sequence and annotation data:

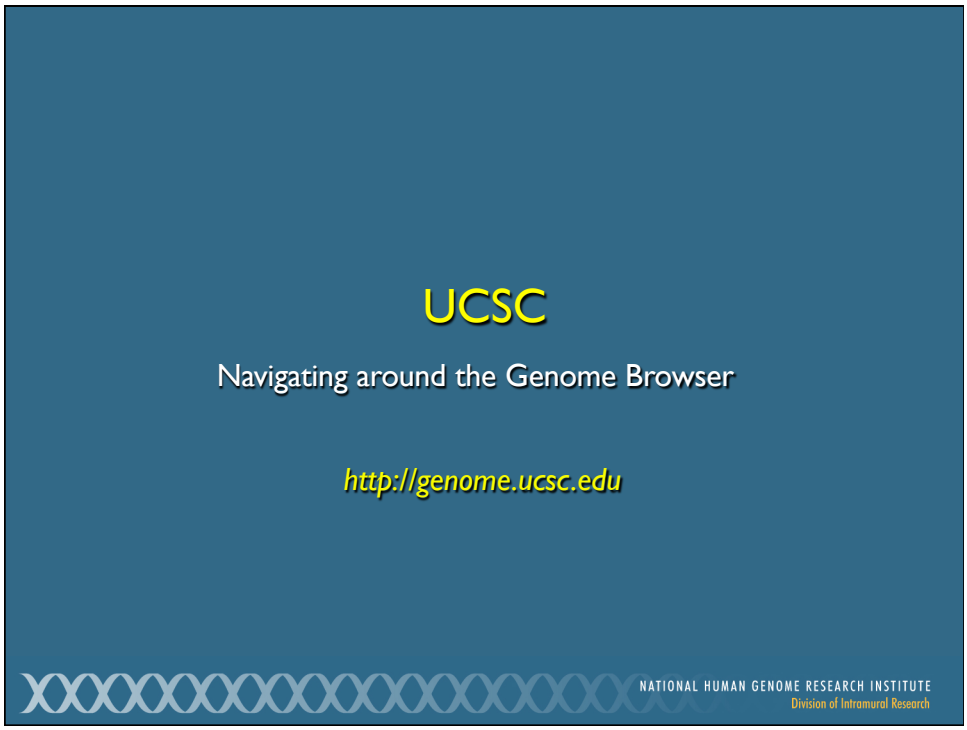

Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
click
exon
intron
Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.
Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
Mapping and Sequencing

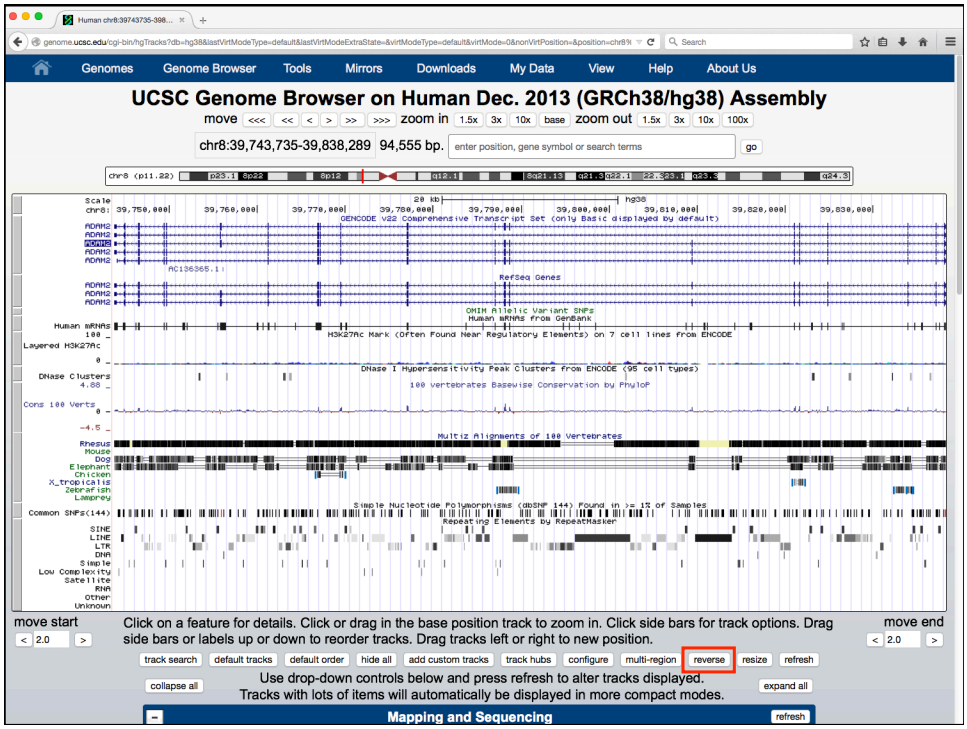

Human Gene ADAM2 (uc003xnk.5) Description and Page Index
UCSC Gene details
Sequence and Links to Tools and Databases
Microarray Expression Data
GNF Expression Atlas 2 Data from U133A and GNF1H Chips
Ratios
Absolute

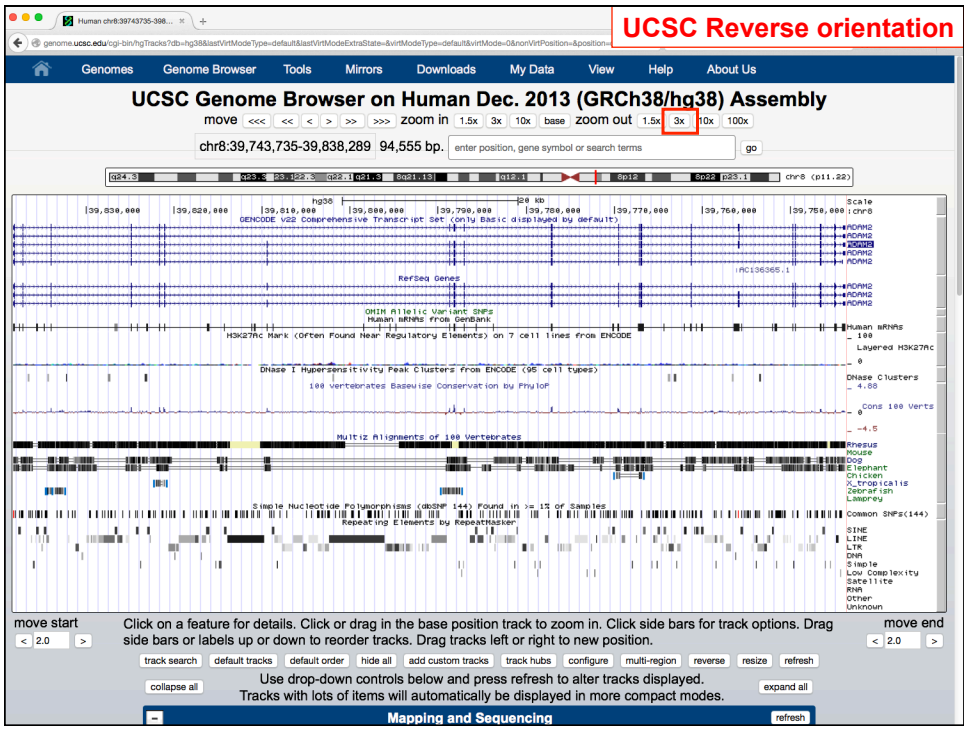

Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
chr8:39,743,735-39,838,289 94,555 bp.
click
RefSeq Genes
Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.
Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
Mapping and Sequencing

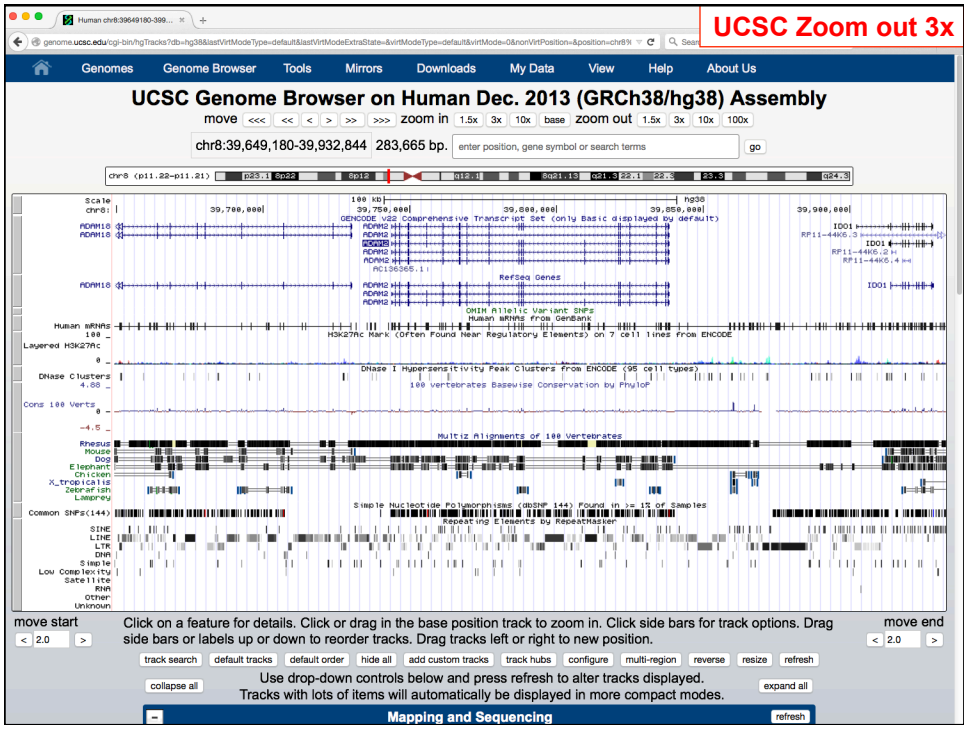

UCSC RefSeq Gene details
RefSeq Gene ADAM2
RefSeq: NM_001464.4 Status: Reviewed
Description: Homo sapiens ADAM metallopeptidase domain 2 (ADAM2), transcript variant 1, mRNA.
CCDS: CCDS34884.1
CDS: 3' complete
OMIM: 601533
Entrez Gene: 2515
PubMed on Gene: ADAM2
PubMed on Product: disintegrin and metalloproteinase 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
The alignment you clicked on is first in the table below.
BROWSER | SIZE IDENTITY CHROMOSOME STRAND START END QUERY START END TOTAL
browser | 2671 100.0% 8 - 39743736 39838289 NM_001464 1 2671 2686
browser | 2671 100.0% 8 KI270822v1_alt - 494375 588925 NM_001464 1 2671 2686
Links to sequence:
Predicted Protein
mRNA Sequence
Genomic Sequence
CDS FASTA alignment from multiple alignment
click

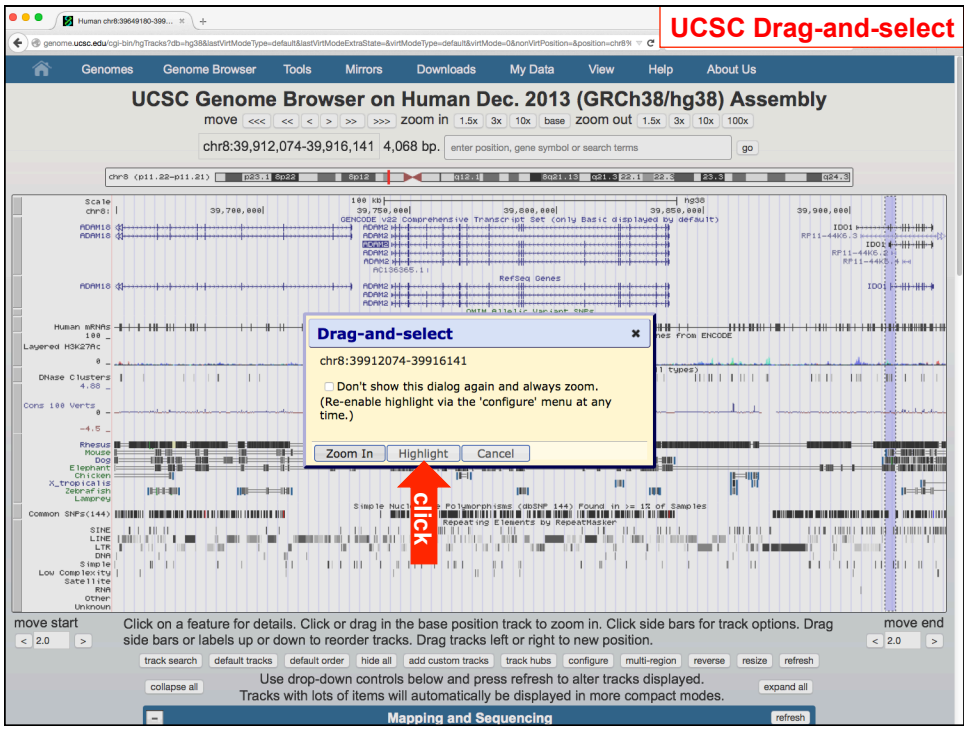

UCSC RefSeq Gene details
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
Help
About Us
Genomic Sequence Near Gene
Get Genomic Sequence Near Gene
Note: if you would prefer to get DNA for more than one feature of this track at a time, try the Table Browser using the output format sequence.
Sequence 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 FASTA r
Note: if a feature is close to the beginning or end of a chromoson
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
>hg38_refGene_NM_001464 range=chr8:39838290-39839289 5'pad=0 3'pad=0
taaattagtacagtcttggaaaacttgtcggaagtatctaccaacacata
ccctgtgatccgacaacttcactctagaaatatacacagtagaaatcctt
acttattacaccaaaaggcatgtagaagaatgtttatagctaaatattat
ttttaatagctggaacaataaacaaacaaaatattcattaacagtaaaat
ggaaacacaagtggtgttatatttaatgaatttgtaatatacaccaatga
ggataaaacagaactattgctttgtatgaaccttacaatcatcctattaa
aagaacccagacatgaaagagtagatggtgattgcttctacttgcgaaaa
gttcaaaaacagacaaaacgaatctttggttgtttagaagtcatggttgt
gaggttggaatctggggatttgggtggttctttttcatttcttcatcgtg
gtactagttacgctgtttttttttttccacttgaatattaattgacctgtg
aacttatgattatctatatactttttttcgttttttgtttcgttctttttt
ttttttttttttttttttttttgacggaatttcgctctctcacccaggctg
gagtgcagtggtaaggtctttctgctcactgcaacctctgcctcctaggt
tcaagcgattctcctgcctcagcttcccgagtagctgggatttcaggcac
ccgccaccatgcctggctaatttttttttgtatttttagtacagacgggg
tttcaccatgttgcggaaactggtctcgaactcctgatacgtgtttatat
atattcaattgaaatttacttaagaagtgtttataaatcttcctgttcc
tcagctgttggaagtgattttgttgttgctgttgccttaattaggcatca
cgtccagtggagtgctgtcctgcaagagacagggctcaggagctcgcac
gttccacagcgcaccacacccaacctcagcccacctgggcttcccagcc

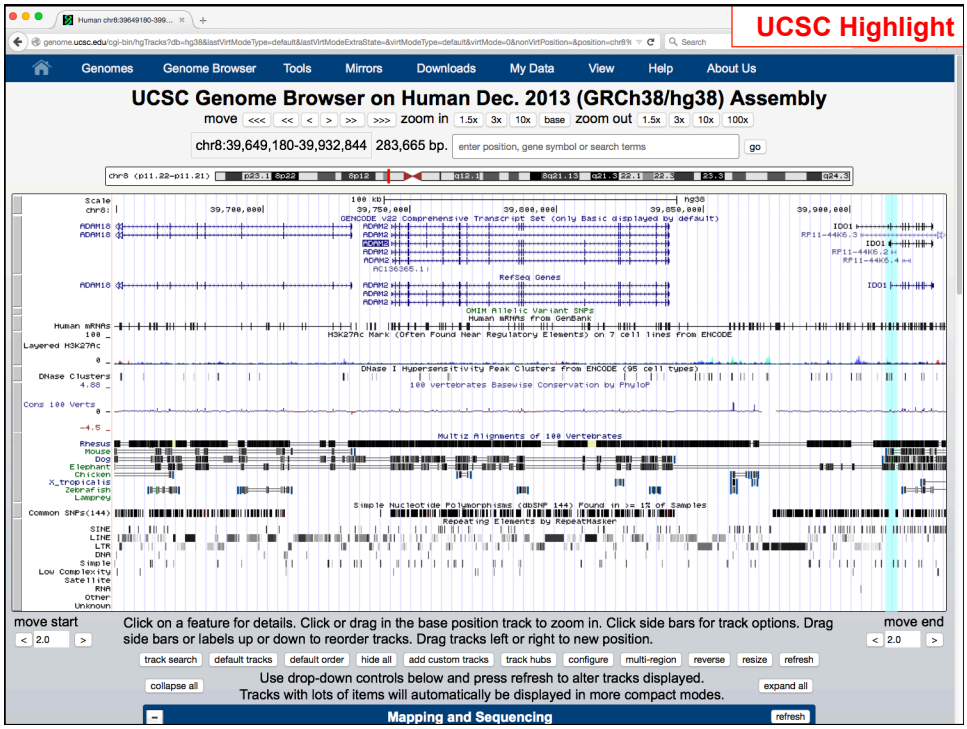

UCSC
Navigating around the Genome Browser
http://genome.ucsc.edu
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

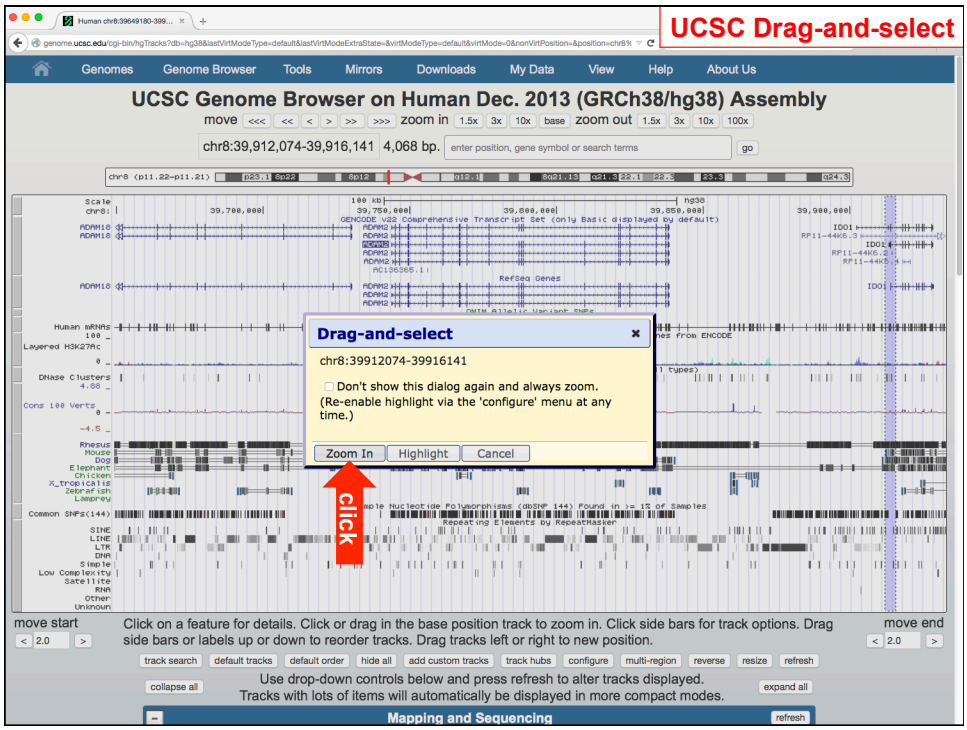

Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
move
zoom in
zoom out
chr8:39,743,735-39,838,289 94,555 bp.
enter position, gene symbol or search terms
go
GENCODE V22 Comprehensive Transcript Set (only Basic displayed by default)
RefSeq Genes
OMIM Allelic Variant SNPs
Human mRNAs from GenBank
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
DNase I Hypersensitivity Peak Clusters from ENCODE (95 cell types)
100 Vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 144) 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
multi-region
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

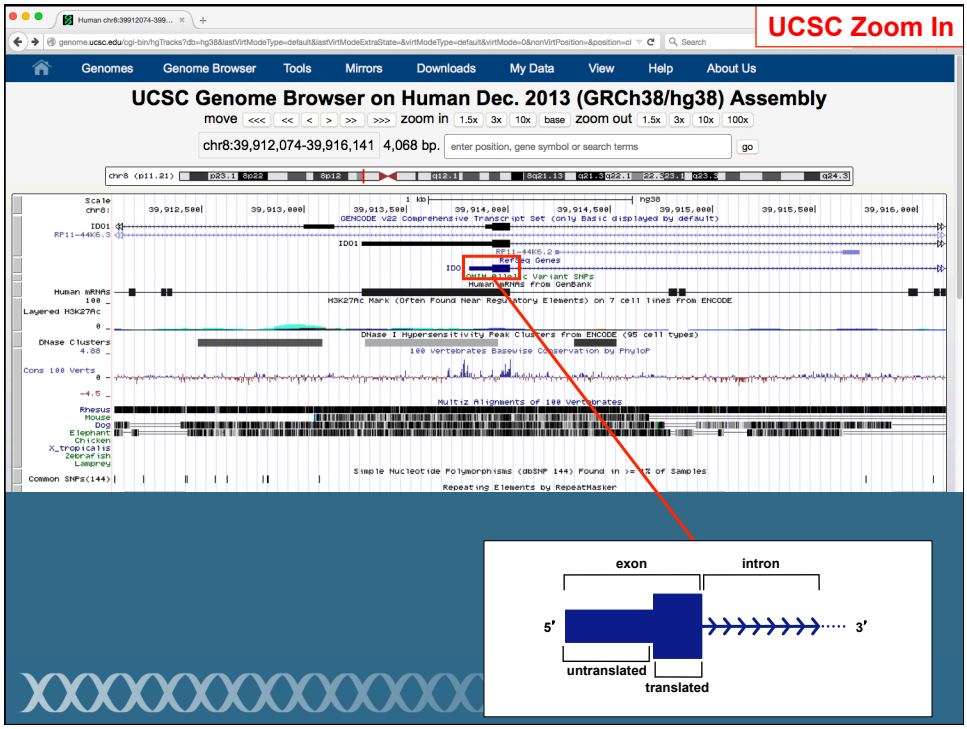

UCSC Reverse orientation
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
move
zoom in
zoom out
chr8:39,743,735-39,838,289 94,555 bp.
enter position, gene symbol or search terms
go
GENCODE V22 Comprehensive Transcript Set (only Basic displayed by default)
RefSeq Genes
OMIM Allelic Variant SNPs
Human mRNAs from GenBank
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
DNase I Hypersensitivity Peak Clusters from ENCODE (95 cell types)
100 Vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 144) 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
multi-region
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

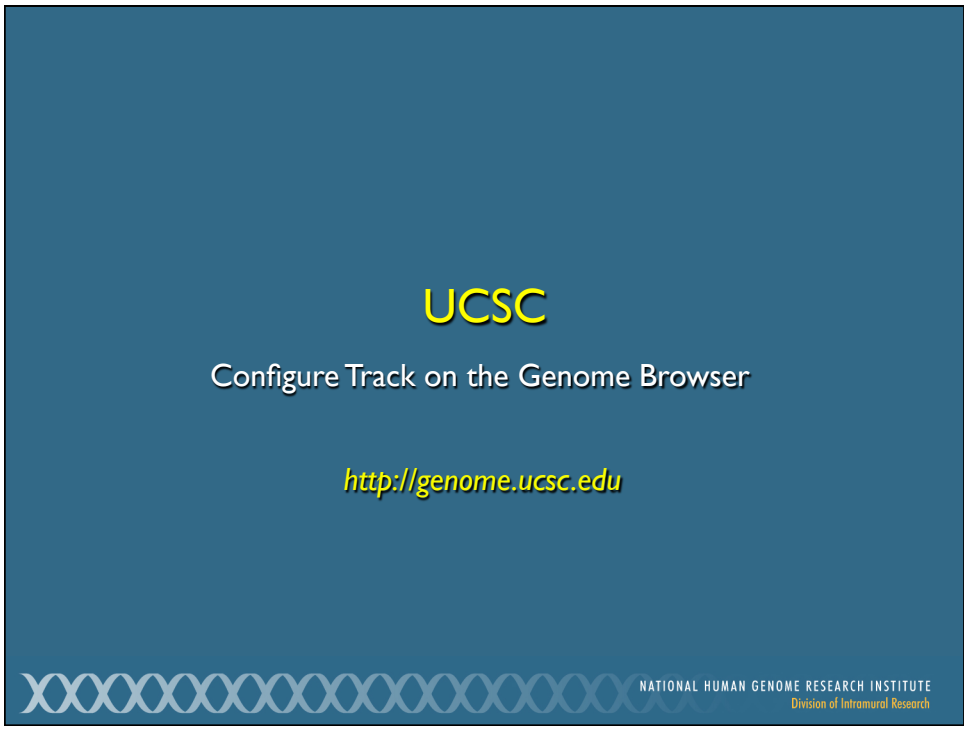

UCSC Zoom out 3x
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
move
zoom in
zoom out
chr8:39,649,180-39,932,844
283,665 bp.
enter position, gene symbol or search terms
go
move start
move end
Click on a feature for 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
multi-region
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

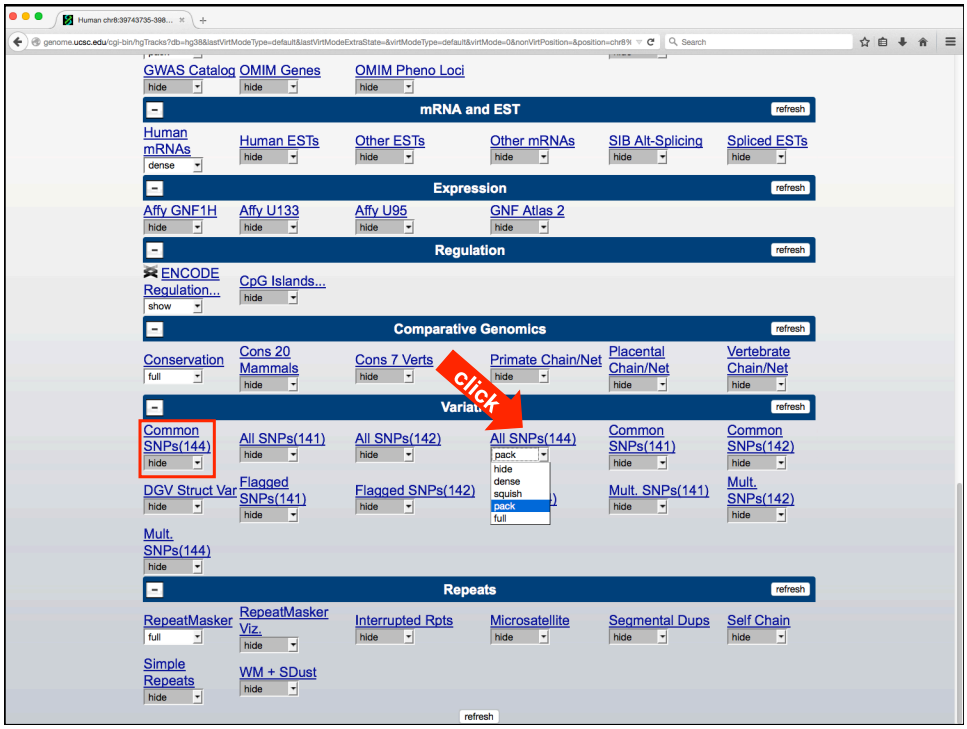

UCSC Drag-and-select
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
move
zoom in
zoom out
chr8:39,912,074-39,916,141
4,068 bp.
enter position, gene symbol or search terms
go
Drag-and-select
chr8:39912074-39916141
Don't show this dialog again and always zoom.
(Re-enable highlight via the 'configure' menu at any time.)
Zoom In
Highlight
Cancel
click
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
multi-region
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

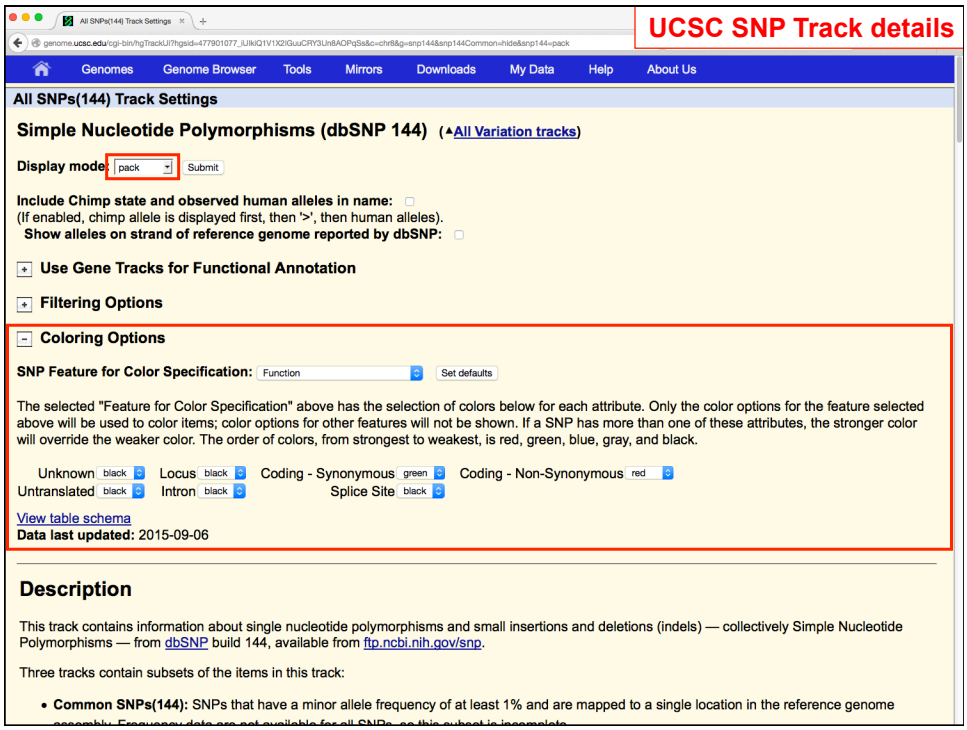

UCSC Highlight
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
chr8:39,649,180-39,932,844 283,665 bp.
move start
move end
Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.
Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
Mapping and Sequencing

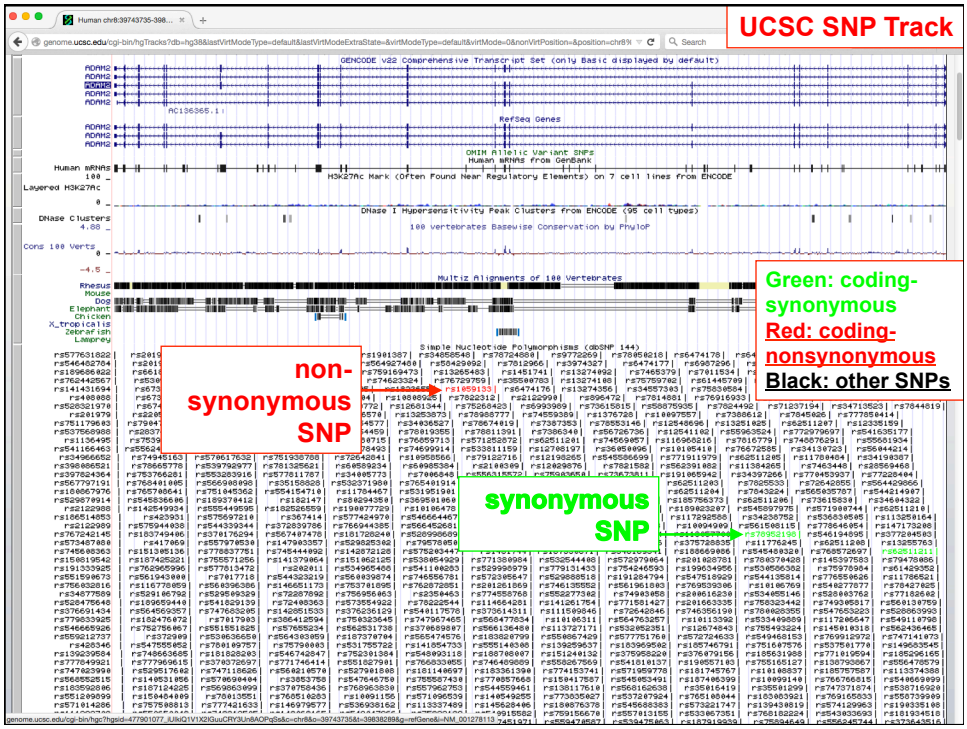

UCSC Drag-and-select
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
chr8:39,912,074-39,916,141 4,068 bp.
Drag-and-select
chr8:39912074-39916141
Don't show this dialog again and always zoom.
(Re-enable highlight via the 'configure' menu at any time.)
Zoom In
Highlight
Cancel
click
move start
move end
Mapping and Sequencing

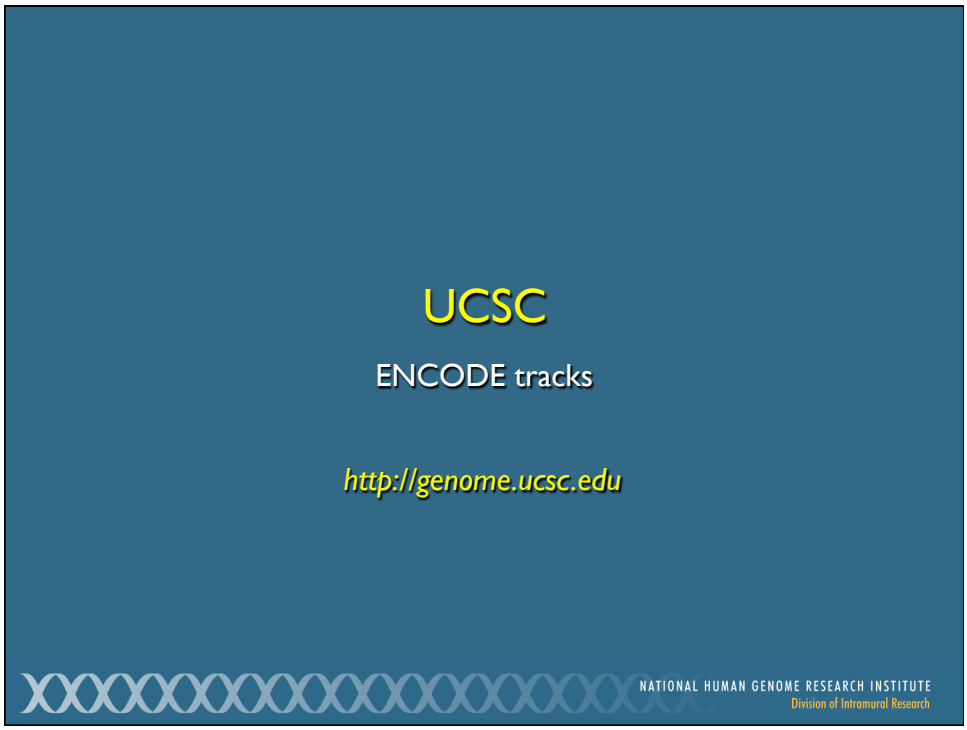

UCSC Zoom In
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
move
zoom in
zoom out
chr8:39,912,074-39,916,141 4,068 bp.
enter position, gene symbol or search terms
go
exon
intron
5'
3'
untranslated
translated

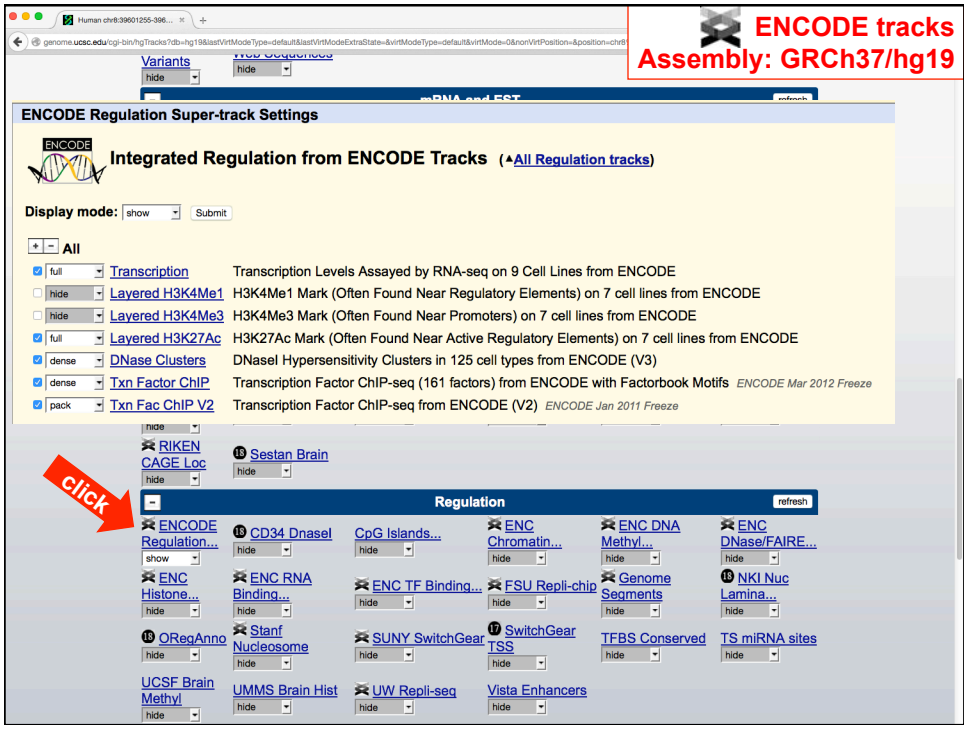

UCSC
Configure Track on the Genome Browser
http://genome.ucsc.edu
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

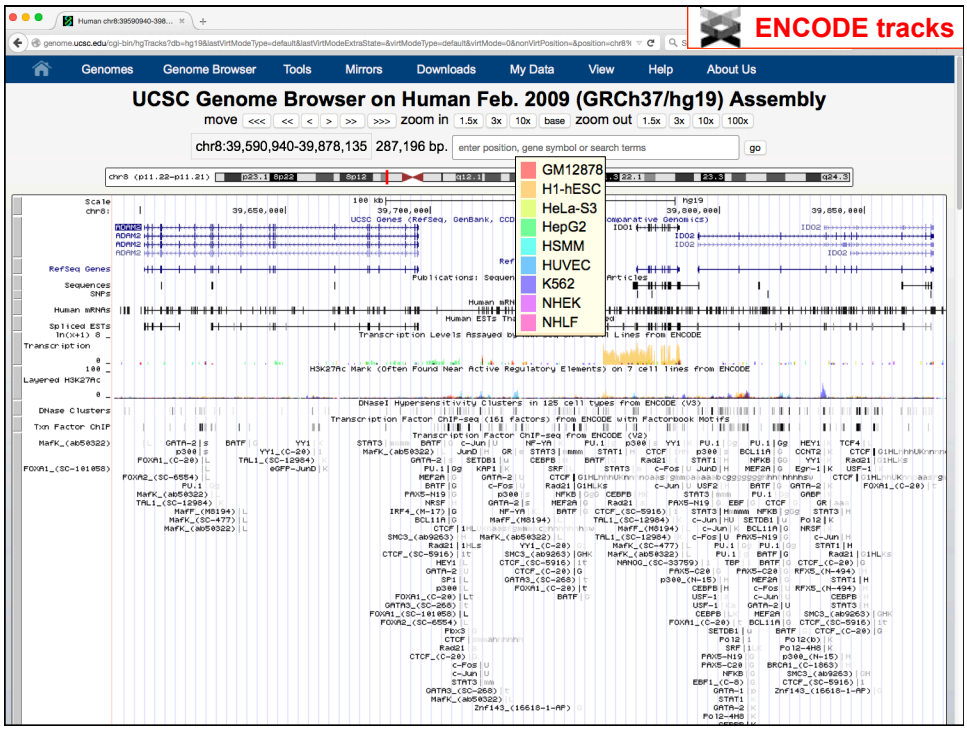

GWAS Catalog
OMIM Genes
OMIM Pheno Loci
mRNA and EST
Human mRNAs
Human ESTs
Other ESTs
Other mRNAs
SIB Alt-Splicing
Spliced ESTs
Expression
Affy GNF1H
Affy U133
Affy U95
GNF Atlas 2
Regulation
ENCODE Regulation...
CpG Islands...
Comparative Genomics
Conservation
Cons 20 Mammals
Cons 7 Verts
Primate Chain/Net
Placental Chain/Net
Vertebrate Chain/Net
click
Common SNPs(144)
All SNPs(141)
All SNPs(142)
All SNPs(144)
Common SNPs(141)
Common SNPs(142)
DGV Struct Var
Flagged SNPs(141)
Flagged SNPs(142)
Mult. SNPs(141)
Mult. SNPs(142)
Mult. SNPs(144)
hide
dense
squish
pack
full
Repeats
RepeatMasker
RepeatMasker Viz.
Interrupted Rpts
Microsatellite
Segmental Dups
Self Chain
Simple Repeats
WM + SDust
refresh

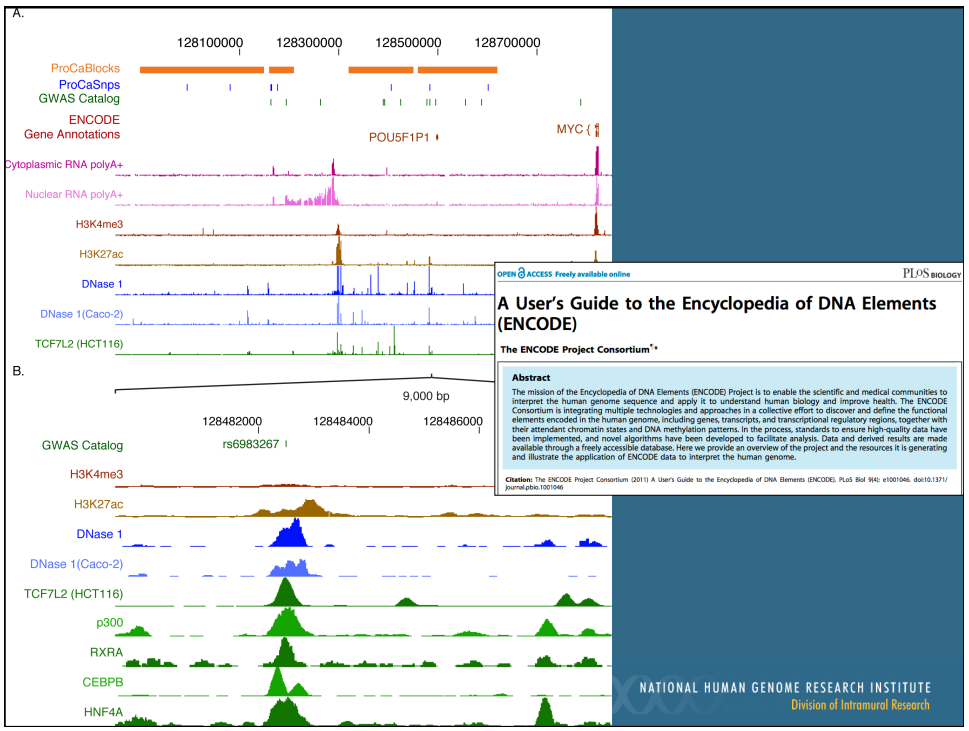

UCSC SNP Track details
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
Help
About Us
All SNPs(144) Track Settings
Simple Nucleotide Polymorphisms (dbSNP 144)
All Variation tracks
Display mode:
pack
Submit
Include Chimp state and observed human alleles in name:
(If enabled, chimp allele is displayed first, then '>', then human alleles).
Show alleles on strand of reference genome reported by dbSNP:
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: 2015-09-06
Description
This track contains information about single nucleotide polymorphisms and small insertions and deletions (indels) — collectively Simple Nucleotide Polymorphisms — from dbSNP build 144, available from ftp.ncbi.nih.gov/snp.
Three tracks contain subsets of the items in this track:
Common SNPs(144): SNPs that have a minor allele frequency of at least 1% and are mapped to a single location in the reference genome


UCSC SNP Track
Green: coding-synonymous
Red: coding-nonsynonymous
Black: other SNPs
non-synonymous SNP
synonymous SNP

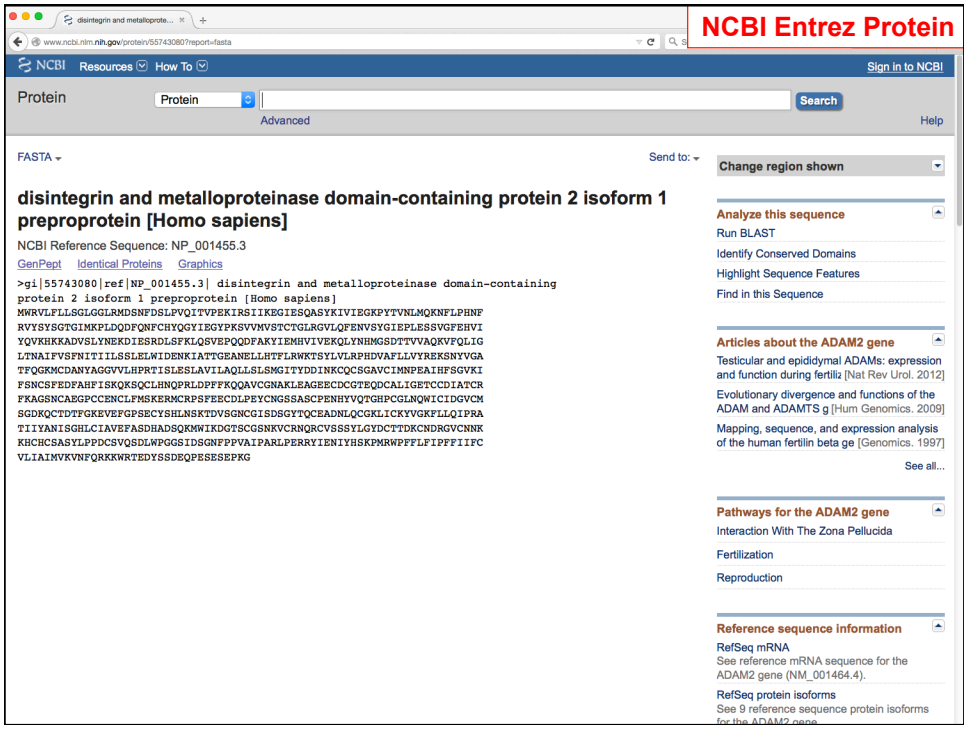

UCSC
ENCODE tracks
http://genome.ucsc.edu
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

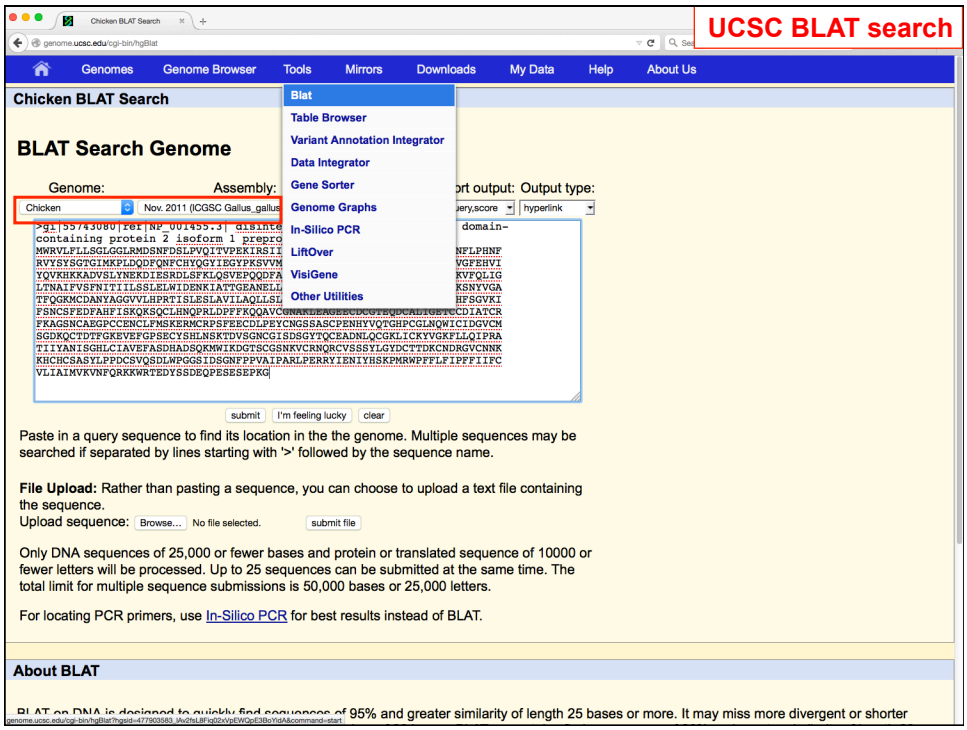

ENCODE tracks
Assembly: GRCh37/hg19
ENCODE Regulation Super-track Settings
Integrated Regulation from ENCODE Tracks
Display mode:
Transcription
Transcription Levels Assayed by RNA-seq on 9 Cell Lines from ENCODE
Layered H3K4Me1
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K4Me3
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
Layered H3K27Ac
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
DNase Clusters
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
Txn Factor ChIP
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
Txn Fac ChIP V2
Transcription Factor ChIP-seq from ENCODE (V2)
click
Regulation
ENCODE Regulation...
CD34 DnaseI
CpG Islands...
ENC Chromatin...
ENC DNA Methyl...
ENC DNase/FAIRE...
ENC Histone...
ENC RNA Binding...
ENC TF Binding...
FSU Repli-chip
Genome Segments
NKI Nuc Lamina...
ORegAnno
Stanf Nucleosome
SUNY SwitchGear
SwitchGear TSS
TFBS Conserved
TS miRNA sites
UCSF Brain Methyl
UMMS Brain Hist
UW Repli-seq
Vista Enhancers

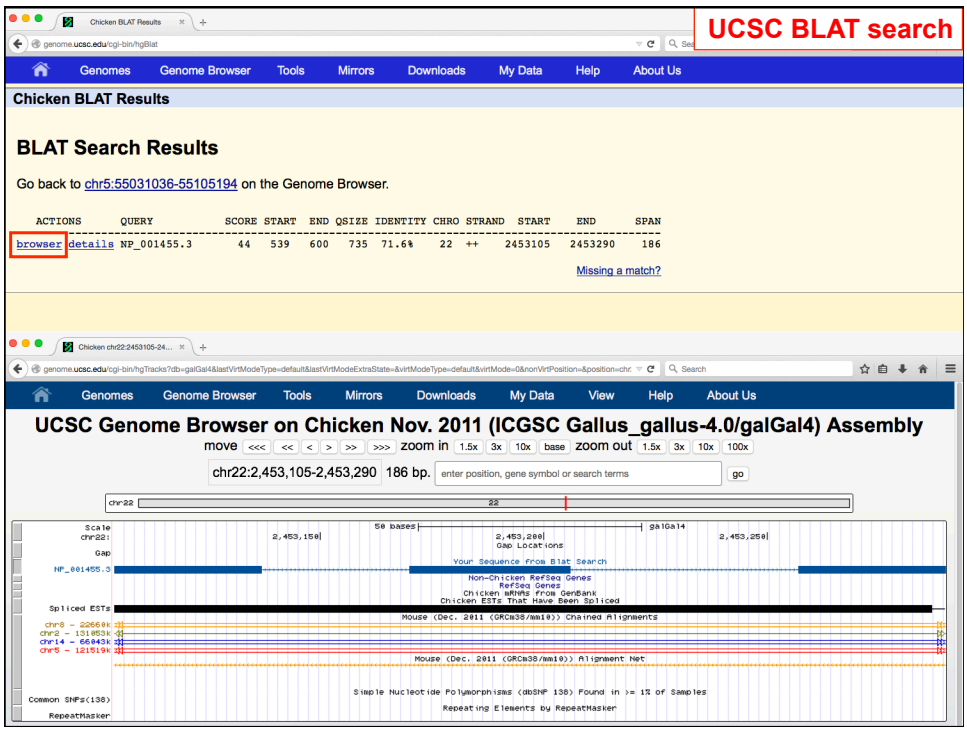

ENCODE tracks
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move
zoom in
zoom out
chr8:39,590,940-39,878,135
287,196 bp.
enter position, gene symbol or search terms
go
GM12878
H1-hESC
HeLa-S3
HepG2
HSMM
HUVEC
K562
NHEK
NHLF
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs
Transcription Factor ChIP-seq from ENCODE (V2)

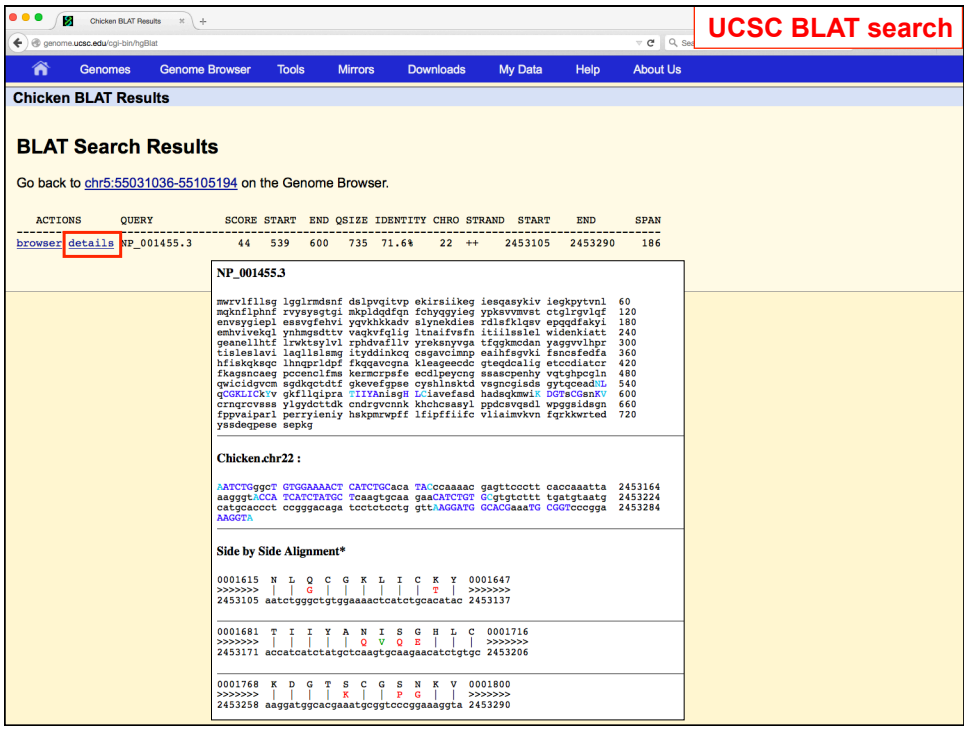

A.
128100000
128300000
128500000
128700000
ProCaBlocks
ProCaSnps
GWAS Catalog
ENCODE
Gene Annotations
POU5F1P1
MYC
Cytoplasmic RNA polyA+
Nuclear RNA polyA+
H3K4me3
H3K27ac
DNase 1
DNase 1(Caco-2)
TCF7L2 (HCT116)
B.
9,000 bp
128482000
128484000
128486000
GWAS Catalog
rs6983267
H3K4me3
H3K27ac
DNase 1
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
The mission of the Encyclopedia of DNA Elements (ENCODE) Project is to enable the scientific and medical communities to interpret the human genome sequence and apply it to understand human biology and improve health. The ENCODE Consortium is integrating multiple technologies and approaches in a collective effort to discover and define the functional elements encoded in the human genome, including genes, transcripts, and transcriptional regulatory regions, together with their attendant chromatin states and DNA methylation patterns. In the process, standards to ensure high-quality data have been implemented, and novel algorithms have been developed to facilitate analysis. Data and derived results are made available through a freely accessible database. Here we provide an overview of the project and the resources it is generating and illustrate the application of ENCODE data to interpret the human genome.
Citation: The ENCODE Project Consortium (2011) A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLoS Biol 9(4): e1001046. doi:10.1371/journal.pbio.1001046
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


NCBI Entrez Protein
NCBI Resources How To
Sign in to NCBI
Protein
Search
Advanced
Help
FASTA
Send to:
Change region shown
disintegrin and metalloproteinase domain-containing protein 2 isoform 1 preproprotein [Homo sapiens]
NCBI Reference Sequence: NP_001455.3
GenPept Identical Proteins Graphics
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 during fertili [Nat Rev Urol. 2012]
Evolutionary divergence and functions of the ADAM and ADAMTS g [Hum Genomics. 2009]
Mapping, sequence, and expression analysis of the human fertilin beta ge [Genomics. 1997]
See all...
Pathways for the ADAM2 gene
Interaction With The Zona Pellucida
Fertilization
Reproduction
Reference sequence information
RefSeq mRNA
See reference mRNA sequence for the ADAM2 gene (NM_001464.4).
RefSeq protein isoforms


UCSC BLAT search
Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us
Chicken BLAT Search
BLAT Search Genome
Genome:
Assembly:
Chicken
Nov. 2011 (ICGSC Gallus_gallus
Blat
Table Browser
Variant Annotation Integrator
Data Integrator
Gene Sorter
Genome Graphs
In-Silico PCR
LiftOver
VisiGene
Other Utilities
Output type:
hyperlink
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.
About BLAT


UCSC BLAT search
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
Help
About Us
Chicken BLAT Results
BLAT Search Results
Go back to chr5:55031036-55105194 on the Genome Browser.
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
Missing a match?
UCSC Genome Browser on Chicken Nov. 2011 (ICGSC Gallus_gallus-4.0/galGal4) Assembly
chr22:2,453,105-2,453,290 186 bp.
Your Sequence from Blat Search
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


UCSC BLAT search
Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
Help
About Us
Chicken BLAT Results
BLAT Search Results
Go back to chr5:55031036-55105194 on the Genome Browser.
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 :
Side by Side Alignment*

UCSC

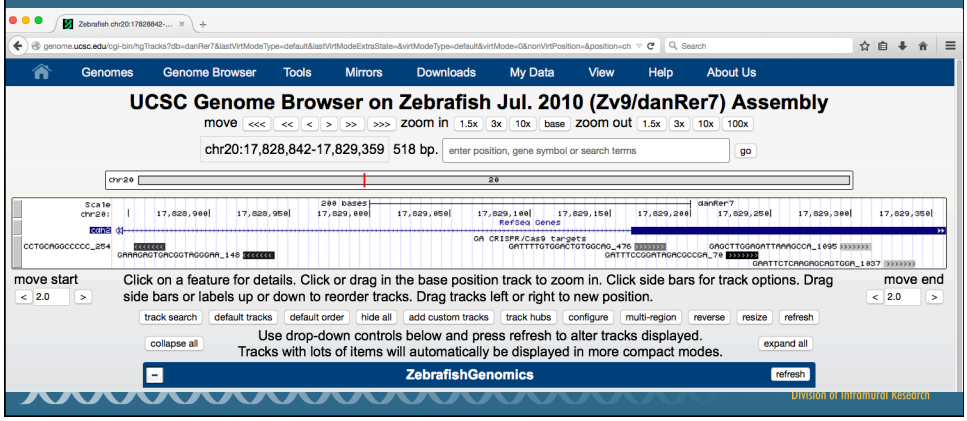
Add your own custom tracks

http://genome.ucsc.edu

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

UCSC custom track format

```
track    name="crispr_sites"     type=bed        useScore=1      db=danRer7
chr20    17828853        17828873        GACATCCCCTGCAGGCCCCC_254        808.143 -
chr20    17828921        17828941        GAAAGAGTGACGGTAGGGAA_148        923.927 -
chr20    17829165        17829185        GATTTTGTGGACTGTGGCAG_476        576.42  +
chr20    17829222        17829242        GATTTCCGGATAGACGCCGA_70         986.378 +
chr20    17829293        17829313        GAGCTTGGAGATTAAAGCCA_1095       216.377 +
chr20    17829320        17829340        GAATTCTCAAGAGCAGTGGA_1037       235.802 +
```



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
  - Session can be shared or non-shared
  - Session persists for 4 months; custom tracks for 48 hours
- Set up a Hub to share very large data sets
  - Hub tracks can be grouped into composite or super-tracks
  - Supports genome assemblies not available at UCSC

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

UCSC

Table Browser

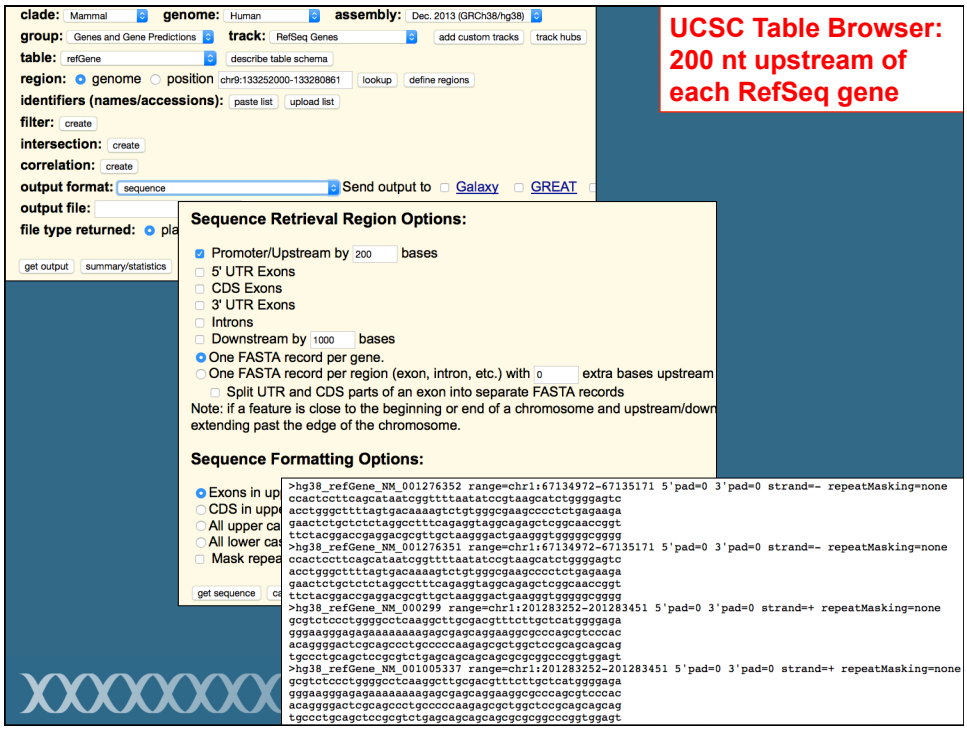
http://genome.ucsc.edu

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



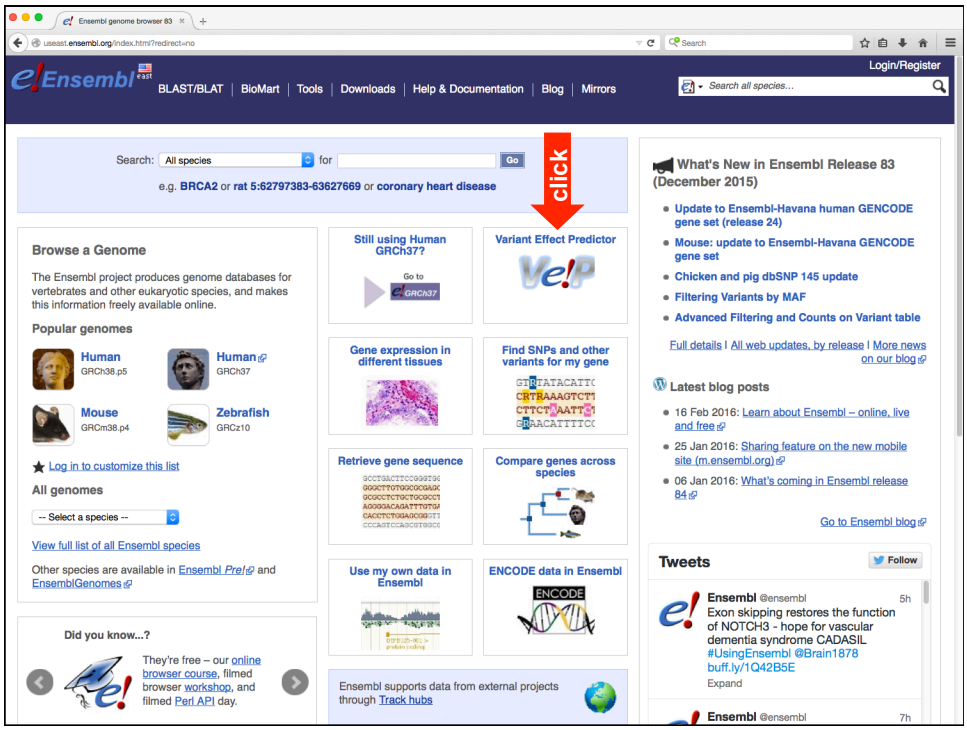
UCSC Table Browser: 200 nt upstream of each RefSeq gene

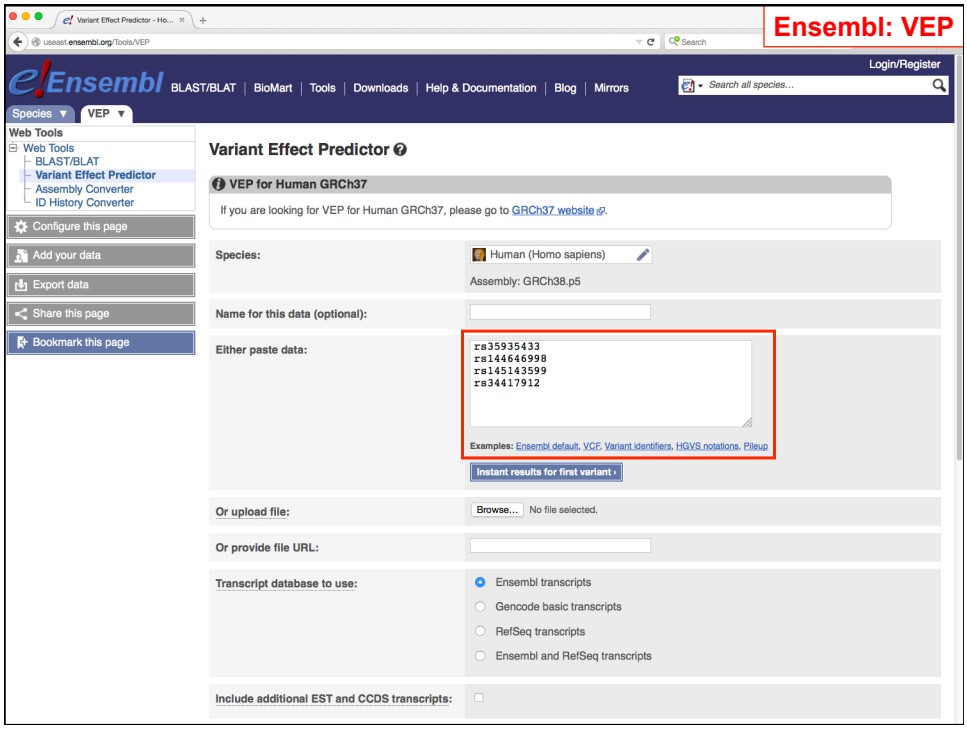
Ensembl

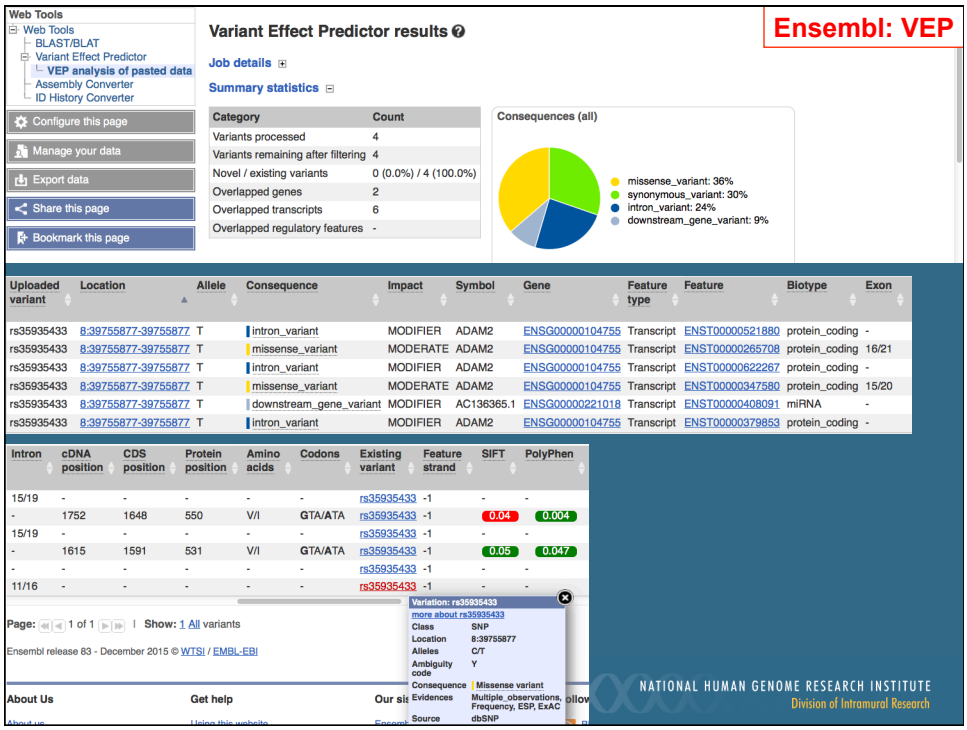
Variant Effect Predictor (VEP)

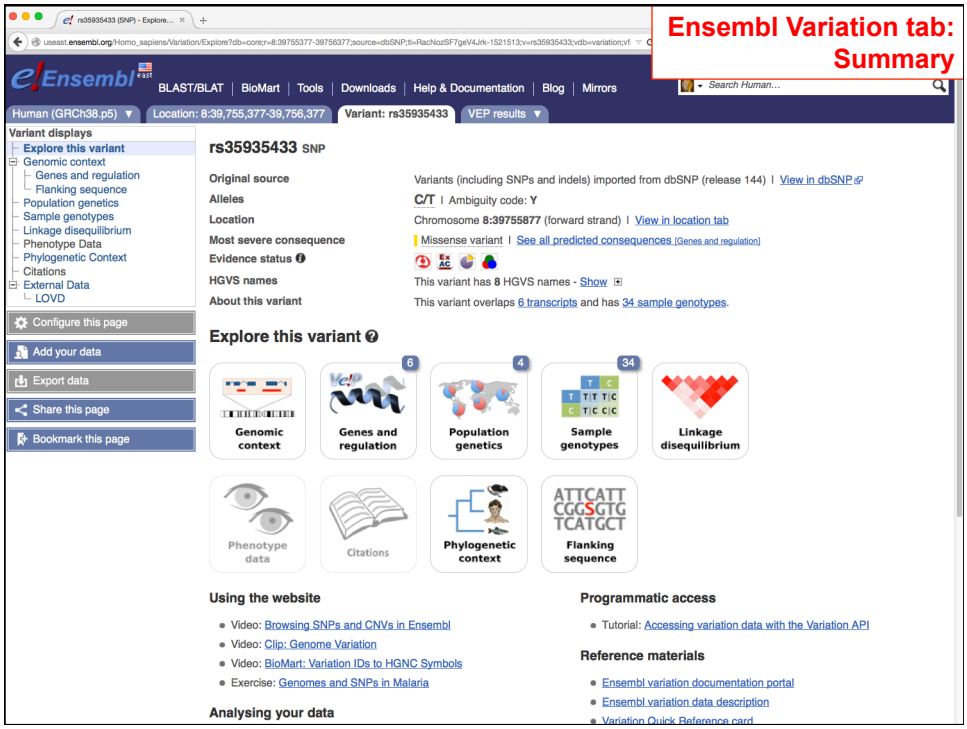
http://www.ensembl.org

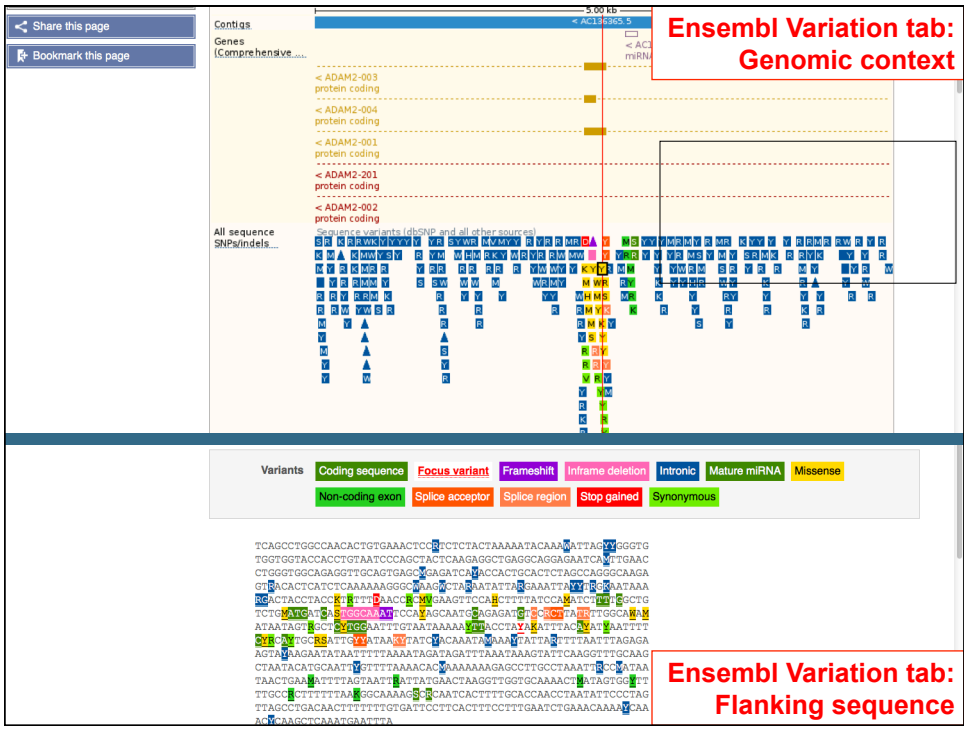
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research




Ensembl: VEP
Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
Species
VEP
Web Tools
Web Tools
BLAST/BLAT
Variant Effect Predictor
Assembly Converter
ID History Converter
Configure this page
Add your data
Export data
Share this page
Bookmark this page
Variant Effect Predictor
VEP for Human GRCh37
If you are looking for VEP for Human GRCh37, please go to GRCh37 website.
Species:
Human (Homo sapiens)
Assembly: GRCh38.p5
Name for this data (optional):
Either paste data:
rs35935433
rs144646998
rs145143599
rs34417912
Examples: Ensembl default, VCF, Variant identifiers, HGVS notations, Pileup
Instant results for first variant
Or upload file:
Browse...
No file selected.
Or provide file URL:
Transcript database to use:
Ensembl transcripts
Gencode basic transcripts
RefSeq transcripts
Ensembl and RefSeq transcripts
Include additional EST and CCDS transcripts:


Ensembl: VEP
Web Tools
Web Tools
BLAST/BLAT
Variant Effect Predictor
VEP analysis of pasted data
Assembly Converter
ID History Converter
Configure this page
Manage your data
Export data
Share this page
Bookmark this page
Variant Effect Predictor results
Job details
Summary statistics
Category
Count
Variants processed
4
Variants remaining after filtering
4
Novel / existing variants
0 (0.0%) / 4 (100.0%)
Overlapped genes
2
Overlapped transcripts
6
Overlapped regulatory features
-
Consequences (all)
missense_variant: 36%
synonymous_variant: 30%
intron_variant: 24%
downstream_gene_variant: 9%
Uploaded variant
Location
Allele
Consequence
Impact
Symbol
Gene
Feature type
Feature
Biotype
Exon
rs35935433 8:39755877-39755877 T intron_variant MODIFIER ADAM2 ENSG00000104755 Transcript ENST00000521880 protein_coding -
rs35935433 8:39755877-39755877 T missense_variant MODERATE ADAM2 ENSG00000104755 Transcript ENST00000265708 protein_coding 16/21
rs35935433 8:39755877-39755877 T intron_variant MODIFIER ADAM2 ENSG00000104755 Transcript ENST00000622267 protein_coding -
rs35935433 8:39755877-39755877 T missense_variant MODERATE ADAM2 ENSG00000104755 Transcript ENST00000347580 protein_coding 15/20
rs35935433 8:39755877-39755877 T downstream_gene_variant MODIFIER AC136365.1 ENSG00000221018 Transcript ENST00000408091 miRNA -
rs35935433 8:39755877-39755877 T intron_variant MODIFIER ADAM2 ENSG00000104755 Transcript ENST00000379853 protein_coding -
Intron
cDNA position
CDS position
Protein position
Amino acids
Codons
Existing variant
Feature strand
SIFT
PolyPhen
15/19 - - - - - rs35935433 -1 - -
- 1752 1648 550 V/I GTA/ATA rs35935433 -1 0.04 0.004
15/19 - - - - - rs35935433 -1 - -
- 1615 1591 531 V/I GTA/ATA rs35935433 -1 0.05 0.047
- - - - - - rs35935433 -1 - -
11/16 - - - - - rs35935433 -1 - -
Variation: rs35935433
more about rs35935433
Class SNP
Location 8:39755877
Alleles C/T
Ambiguity code Y
Consequence Missense variant
Evidences Multiple_observations, Frequency, ESP, ExAC
Source dbSNP
Page: 1 of 1
Show: 1 All variants
Ensembl release 83 - December 2015 © WTSI / EMBL-EBI
About Us
Get help
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research


Ensembl Variation tab:
Summary
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Human (GRCh38.p5)
Location: 8:39,755,377-39,756,377
Variant: rs35935433
VEP results
Variant displays
Explore this variant
Genomic context
Genes and regulation
Flanking sequence
Population genetics
Sample genotypes
Linkage disequilibrium
Phenotype Data
Phylogenetic Context
Citations
External Data
LOVD
Configure this page
Add your data
Export data
Share this page
Bookmark this page
rs35935433 SNP
Original source
Variants (including SNPs and indels) imported from dbSNP (release 144) | View in dbSNP
Alleles
C/T | Ambiguity code: Y
Location
Chromosome 8:39755877 (forward strand) | View in location tab
Most severe consequence
Missense variant | See all predicted consequences [Genes and regulation]
Evidence status
HGVS names
This variant has 8 HGVS names - Show
About this variant
This variant overlaps 6 transcripts and has 34 sample genotypes.
Explore this variant
Genomic context
Genes and regulation
Population genetics
Sample 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

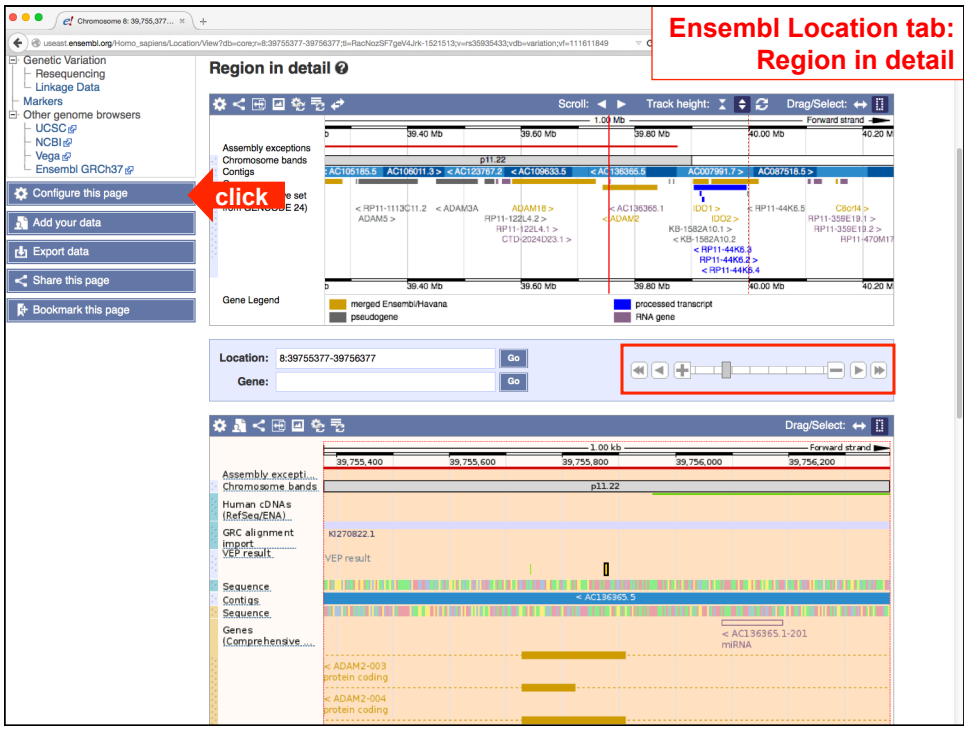

Ensembl Variation tab:
Genomic context
Share this page
Bookmark this page
Contigs
Genes (Comprehensive ...
< ADAM2-003
protein coding
< ADAM2-004
protein coding
< ADAM2-001
protein coding
< ADAM2-201
protein coding
< ADAM2-002
protein coding
All sequence SNPs/indels
Sequence variants (dbSNP and all other sources)
Variants
Coding sequence
Focus variant
Frameshift
Inframe deletion
Intronic
Mature miRNA
Missense
Non-coding exon
Splice acceptor
Splice region
Stop gained
Synonymous
Ensembl Variation tab:
Flanking sequence

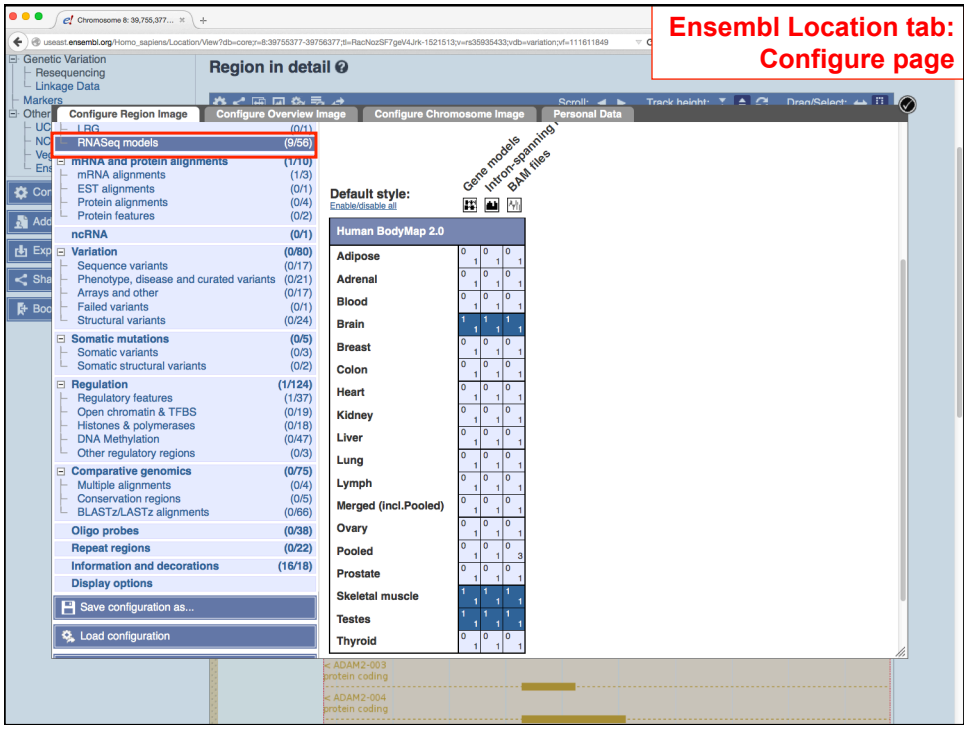
Ensembl

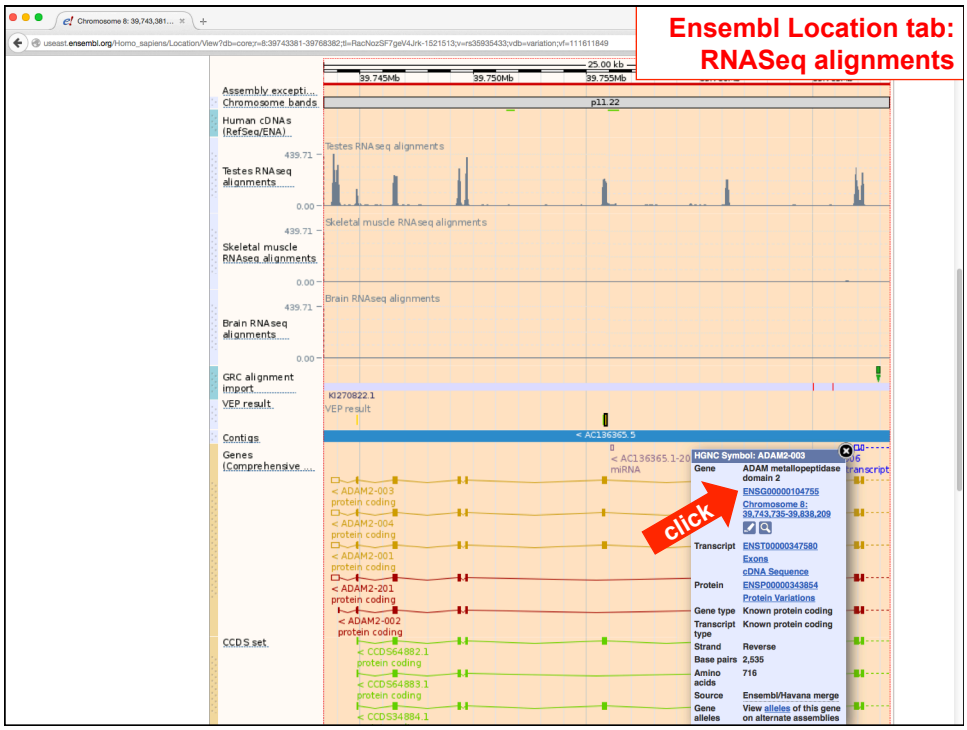
Location tab

http://www.ensembl.org

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research




Ensembl Location tab:
Configure page
Region in detail
Configure Region Image
Configure Overview Image
Configure Chromosome Image
Personal Data
RNASeq models (9/56)
Default style:
Gene models
Intron-spanning
BAM files
Human BodyMap 2.0
Adipose
Adrenal
Blood
Brain
Breast
Colon
Heart
Kidney
Liver
Lung
Lymph
Merged (incl.Pooled)
Ovary
Pooled
Prostate
Skeletal muscle
Testes
Thyroid
Save configuration as...
Load configuration

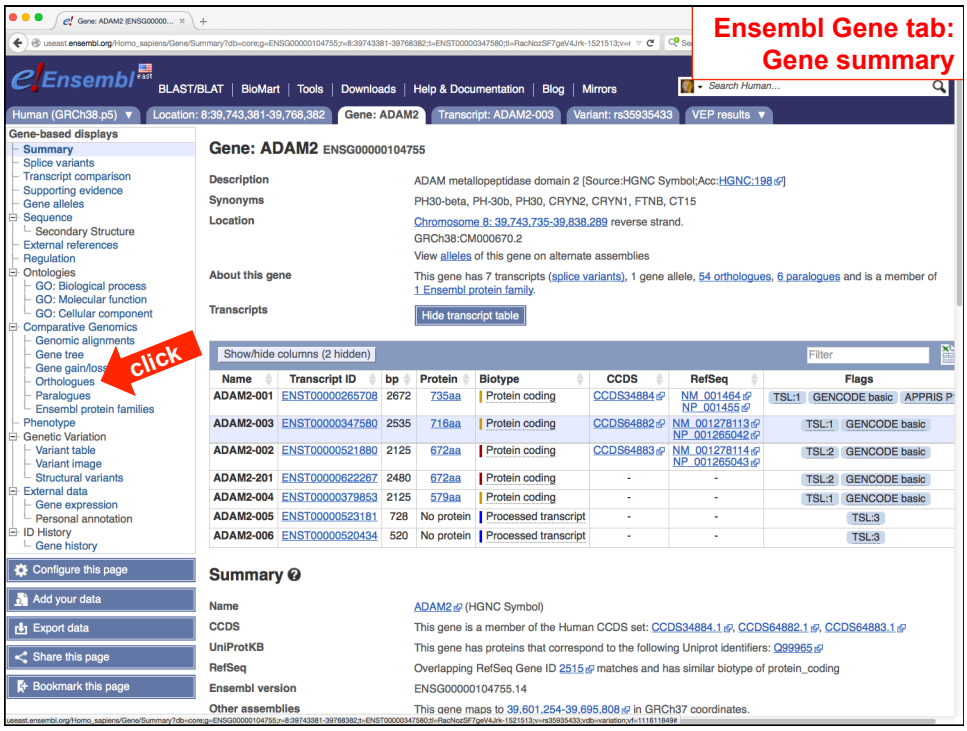

Ensembl Location tab:
RNASeq alignments
Testes RNAseq alignments
Skeletal muscle RNAseq alignments
Brain RNAseq alignments
VEP result
Contigs
Genes (Comprehensive
CCDS set
click
HGNC Symbol: ADAM2-003

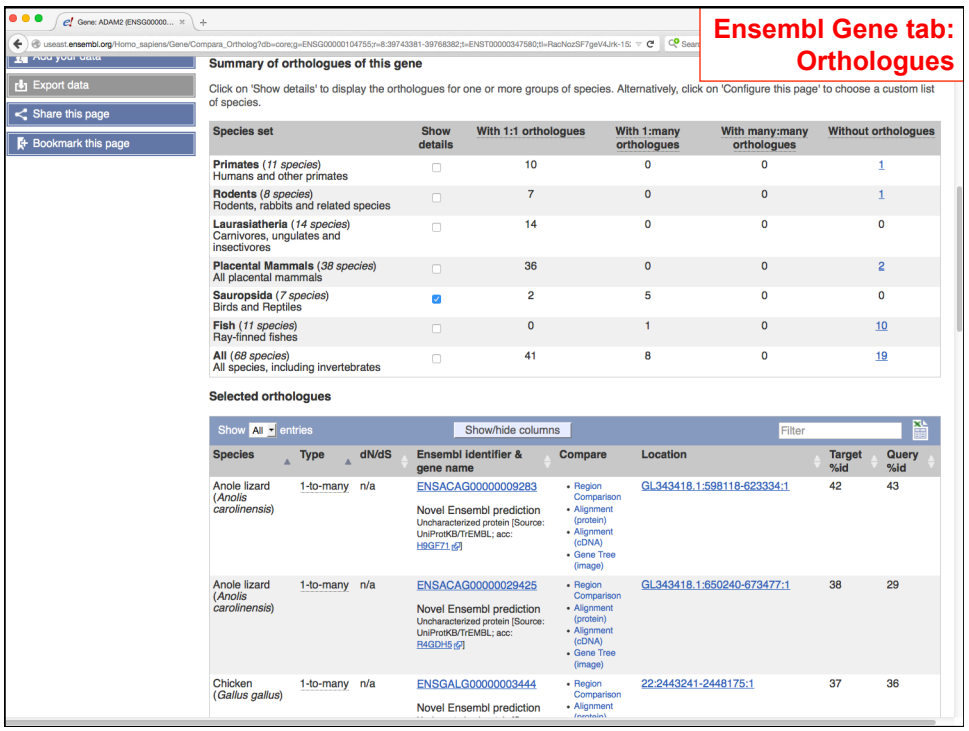
Ensembl

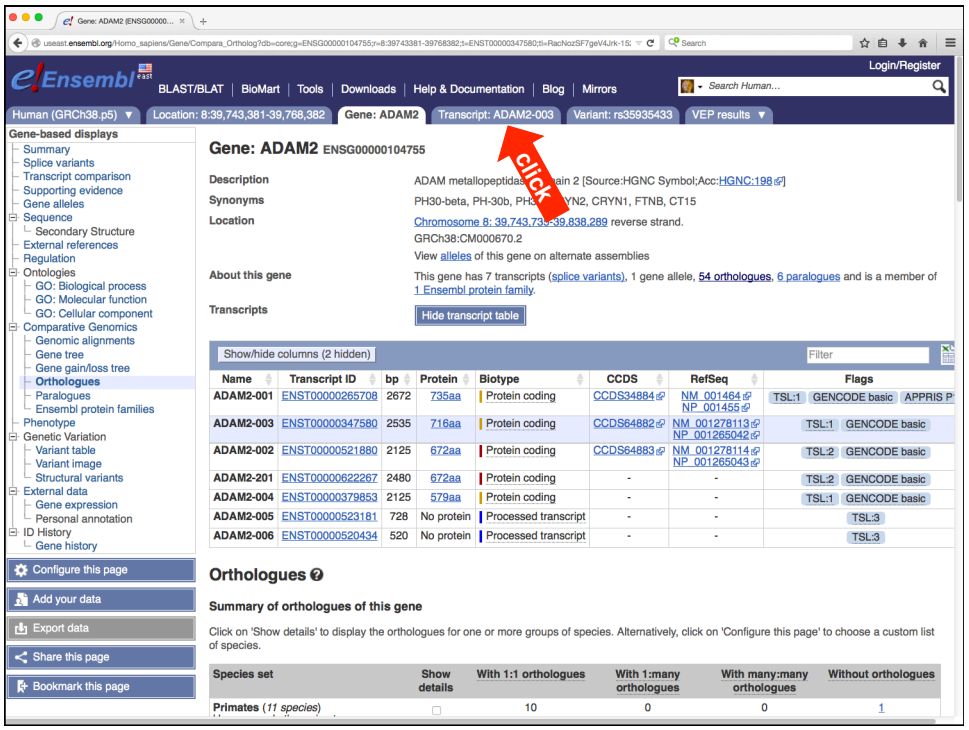
Gene tab

http://www.ensembl.org

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research




Ensembl Gene tab: Orthologues
Export data
Share this page
Bookmark this page
Summary of orthologues of this gene
Click on 'Show details' to display the orthologues for one or more groups of species. Alternatively, click on 'Configure this page' to choose a custom list of species.
Species set | Show details | With 1:1 orthologues | With 1:many orthologues | With many:many orthologues | Without orthologues
Primates (11 species) Humans and other primates | 10 | 0 | 0 | 1
Rodents (8 species) Rodents, rabbits and related species | 7 | 0 | 0 | 1
Laurasiatheria (14 species) Carnivores, ungulates and insectivores | 14 | 0 | 0 | 0
Placental Mammals (38 species) All placental mammals | 36 | 0 | 0 | 2
Sauropsida (7 species) Birds and Reptiles | 2 | 5 | 0 | 0
Fish (11 species) Ray-finned fishes | 0 | 1 | 0 | 10
All (68 species) All species, including invertebrates | 41 | 8 | 0 | 19
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 | Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image) | GL343418.1:598118-623334:1 | 42 | 43
Anole lizard (Anolis carolinensis) | 1-to-many | n/a | ENSACAG00000029425 Novel Ensembl prediction | Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image) | GL343418.1:650240-673477:1 | 38 | 29
Chicken (Gallus gallus) | 1-to-many | n/a | ENSGALG00000003444 Novel Ensembl prediction | Region Comparison, Alignment | 22:2443241-2448175:1 | 37 | 36

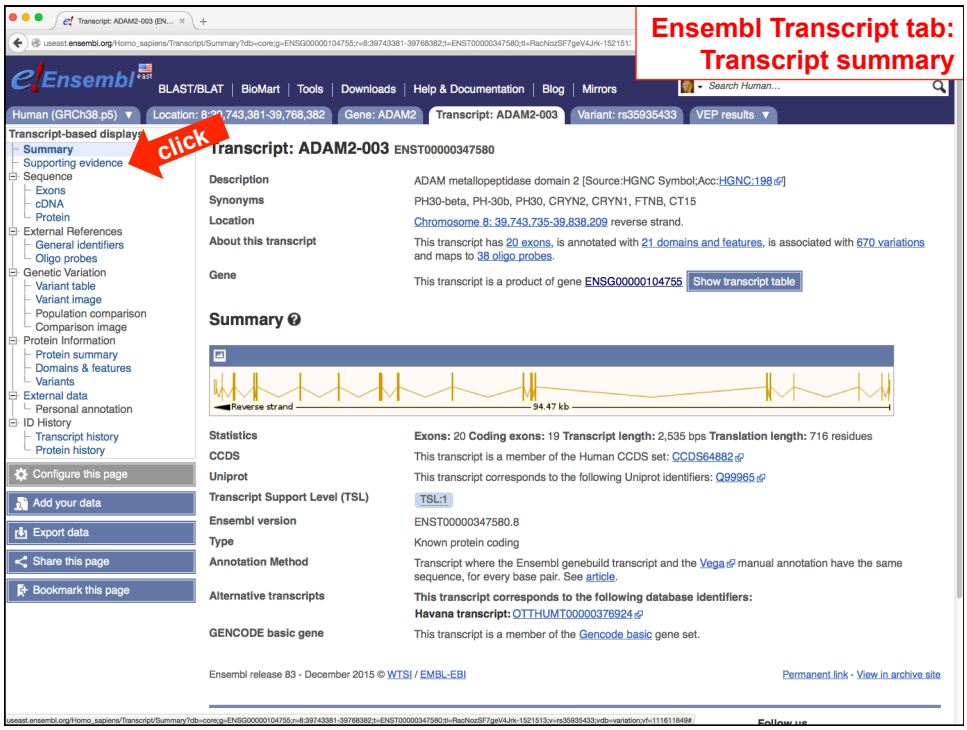

Ensembl
Login/Register
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Search Human...
Human (GRCh38.p5) | Location: 8:39,743,381-39,768,382 | Gene: ADAM2 | Transcript: ADAM2-003 | Variant: rs35935433 | VEP results
click
Gene-based displays
Summary
Splice variants
Transcript comparison
Supporting evidence
Gene alleles
Sequence
Secondary Structure
External references
Regulation
Ontologies
GO: Biological process
GO: Molecular function
GO: Cellular component
Comparative Genomics
Genomic alignments
Gene tree
Gene gain/loss tree
Orthologues
Paralogues
Ensembl protein families
Phenotype
Genetic Variation
Variant table
Variant image
Structural variants
External data
Gene expression
Personal annotation
ID History
Gene history
Gene: ADAM2 ENSG00000104755
Description
Synonyms
PH30-beta, PH-30b, PH
YN2, CRYN1, FTNB, CT15
Location
Chromosome 8: 39,743,73
39,838,289 reverse strand.
GRCh38:CM000670.2
View alleles of this gene on alternate assemblies
About this gene
This gene has 7 transcripts (splice variants), 1 gene allele, 54 orthologues, 6 paralogues and is a member of 1 Ensembl protein family.
Transcripts
Hide transcript table
Show/hide columns (2 hidden)
Filter
Name | Transcript ID | bp | Protein | Biotype | CCDS | RefSeq | Flags
ADAM2-001 | ENST00000265708 | 2672 | 735aa | Protein coding | CCDS34884 | NM_001464 NP_001455 | TSL:1 GENCODE basic APPRIS P
ADAM2-003 | ENST00000347580 | 2535 | 716aa | Protein coding | CCDS64882 | NM_001278113 NP_001265042 | TSL:1 GENCODE basic
ADAM2-002 | ENST00000521880 | 2125 | 672aa | Protein coding | CCDS64883 | NM_001278114 NP_001265043 | TSL:2 GENCODE basic
ADAM2-201 | ENST00000622267 | 2480 | 672aa | Protein coding | - | - | TSL:2 GENCODE basic
ADAM2-004 | ENST00000379853 | 2125 | 579aa | Protein coding | - | - | TSL:1 GENCODE basic
ADAM2-005 | ENST00000523181 | 728 | No protein | Processed transcript | - | - | TSL:3
ADAM2-006 | ENST00000520434 | 520 | No protein | Processed transcript | - | - | TSL:3
Configure this page
Add your data
Export data
Share this page
Bookmark this page
Orthologues
Summary of orthologues of this gene
Click on 'Show details' to display the orthologues for one or more groups of species. Alternatively, click on 'Configure this page' to choose a custom list of species.
Species set | Show details | With 1:1 orthologues | With 1:many orthologues | With many:many orthologues | Without orthologues
Primates (11 species) | 10 | 0 | 0 | 1

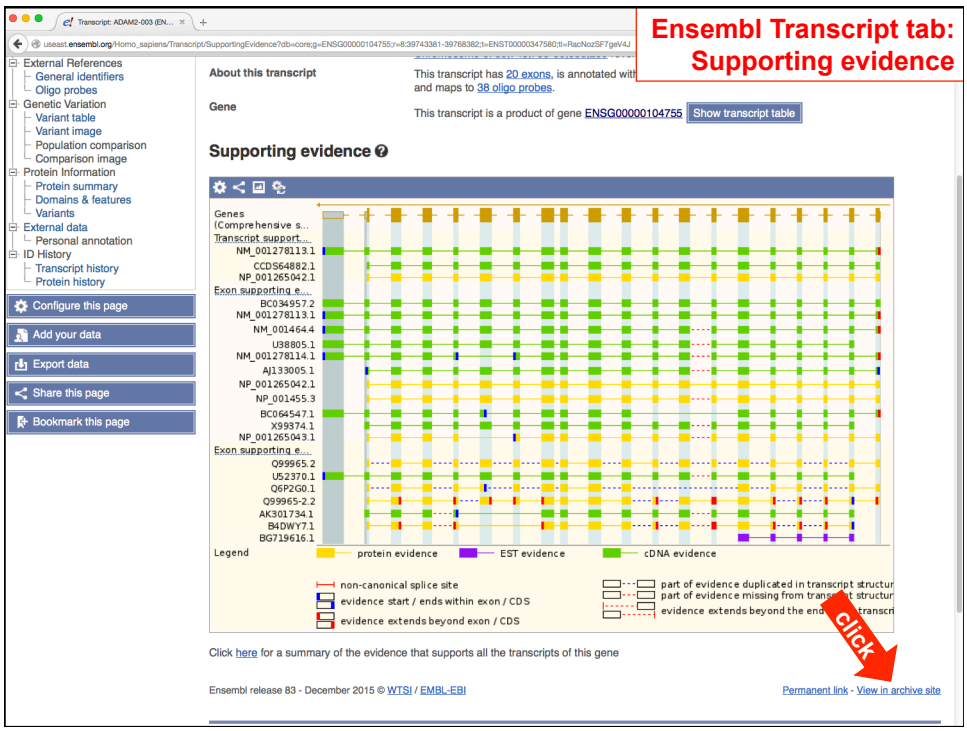
Ensembl

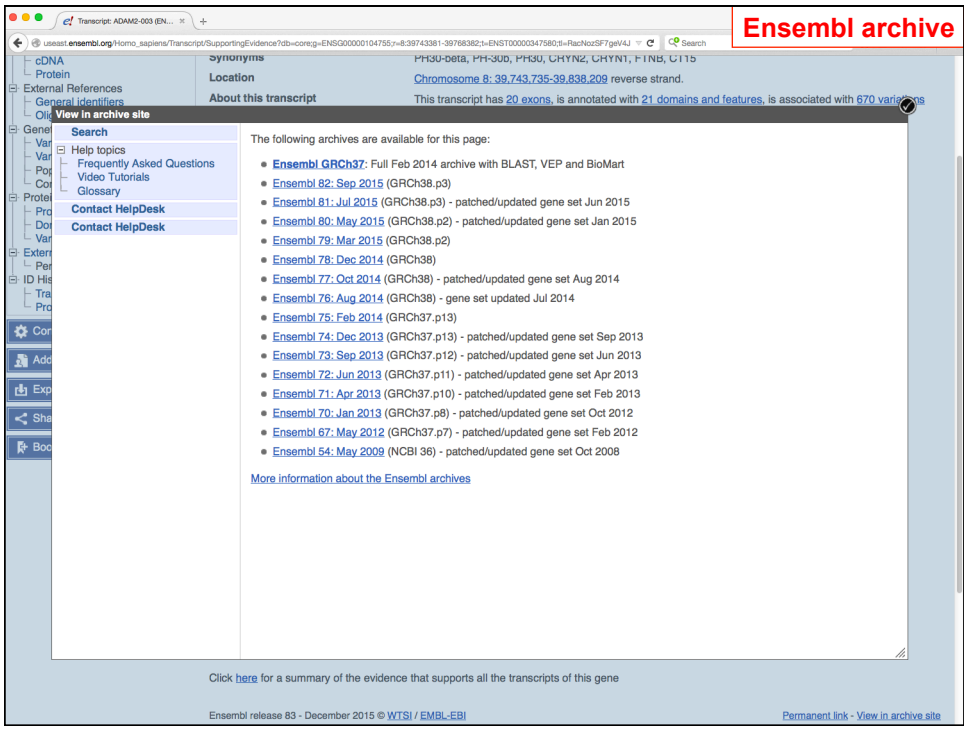
Transcript tab

http://www.ensembl.org

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research




Ensembl Transcript tab:
Supporting evidence
External References
General identifiers
Oligo probes
Genetic Variation
Variant table
Variant image
Population comparison
Comparison image
Protein Information
Protein summary
Domains & features
Variants
External data
Personal annotation
ID History
Transcript history
Protein history
Configure this page
Add your data
Export data
Share this page
Bookmark this page
About this transcript
This transcript has 20 exons, is annotated with
and maps to 38 oligo probes.
Gene
This transcript is a product of gene ENSG00000104755
Show transcript table
Supporting evidence
Genes
(Comprehensive s...
Transcript support...
NM_001278113.1
CCDS64882.1
NP_001265042.1
Exon supporting e...
BC034957.2
NM_001278113.1
NM_001464.4
U38805.1
NM_001278114.1
AJ133005.1
NP_001265042.1
NP_001455.3
BC064547.1
X99374.1
NP_001265043.1
Exon supporting e...
Q99965.2
U52370.1
Q6P2G0.1
Q99965-2.2
AK301734.1
B4DWY7.1
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
click
Click here for a summary of the evidence that supports all the transcripts of this gene
Ensembl release 83 - December 2015 © WTSI / EMBL-EBI
Permanent link - View in archive site

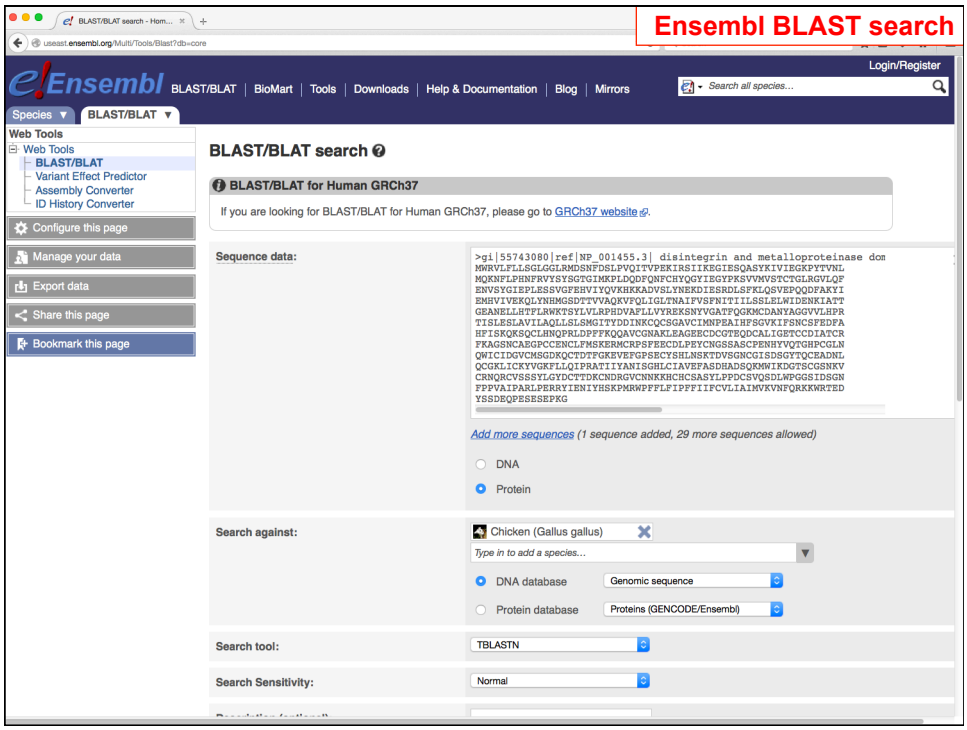

Ensembl archive
cDNA
Protein
External References
General identifiers
Location
Chromosome 8: 39,743,735-39,838,209 reverse strand.
About this transcript
This transcript has 20 exons, is annotated with 21 domains and features, is associated with 670 variations
View in archive site
Search
Help topics
Frequently Asked Questions
Video Tutorials
Glossary
Contact HelpDesk
Contact HelpDesk
The following archives are available for this page:
Ensembl GRCh37: Full Feb 2014 archive with BLAST, VEP and BioMart
Ensembl 82: Sep 2015 (GRCh38.p3)
Ensembl 81: Jul 2015 (GRCh38.p3) - patched/updated gene set Jun 2015
Ensembl 80: May 2015 (GRCh38.p2) - patched/updated gene set Jan 2015
Ensembl 79: Mar 2015 (GRCh38.p2)
Ensembl 78: Dec 2014 (GRCh38)
Ensembl 77: Oct 2014 (GRCh38) - patched/updated gene set Aug 2014
Ensembl 76: Aug 2014 (GRCh38) - gene set updated Jul 2014
Ensembl 75: Feb 2014 (GRCh37.p13)
Ensembl 74: Dec 2013 (GRCh37.p13) - patched/updated gene set Sep 2013
Ensembl 73: Sep 2013 (GRCh37.p12) - patched/updated gene set Jun 2013
Ensembl 72: Jun 2013 (GRCh37.p11) - patched/updated gene set Apr 2013
Ensembl 71: Apr 2013 (GRCh37.p10) - patched/updated gene set Feb 2013
Ensembl 70: Jan 2013 (GRCh37.p8) - patched/updated gene set Oct 2012
Ensembl 67: May 2012 (GRCh37.p7) - patched/updated gene set Feb 2012
Ensembl 54: May 2009 (NCBI 36) - patched/updated gene set Oct 2008
More information about the Ensembl archives
Click here for a summary of the evidence that supports all the transcripts of this gene
Ensembl release 83 - December 2015 © WTSI / EMBL-EBI
Permanent link - View in archive site

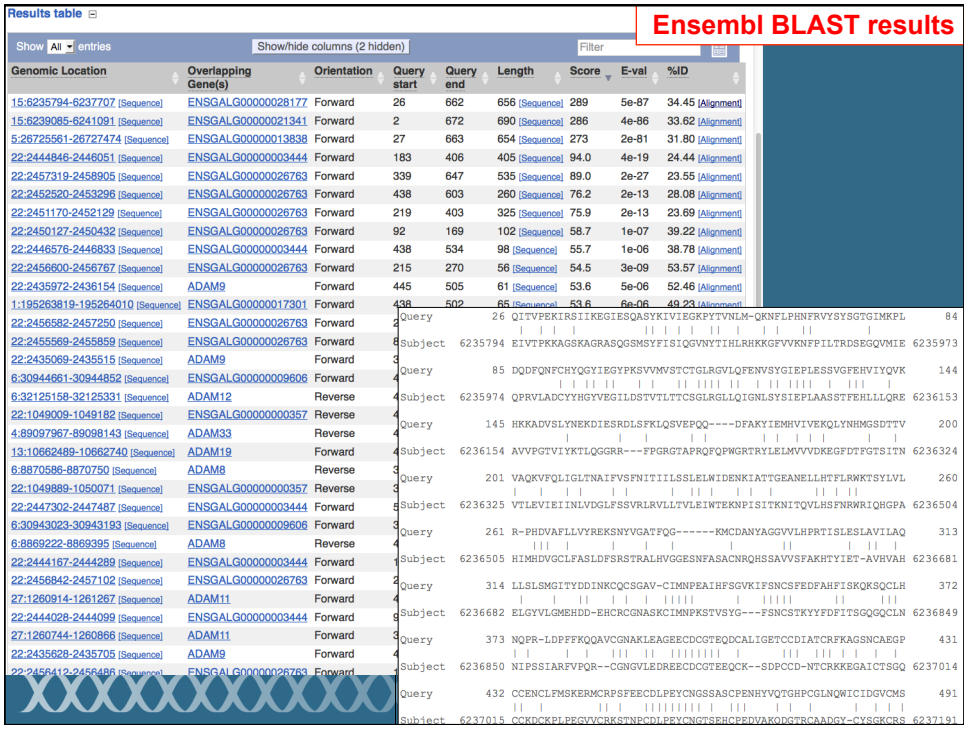
Ensembl

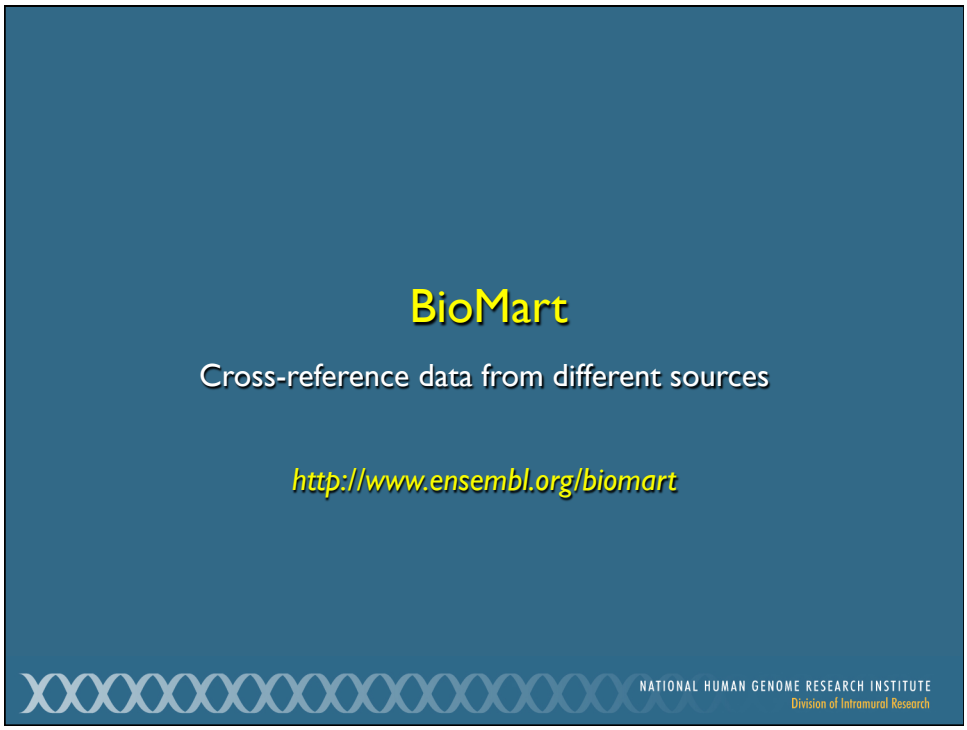
Find a chicken homolog of a human protein

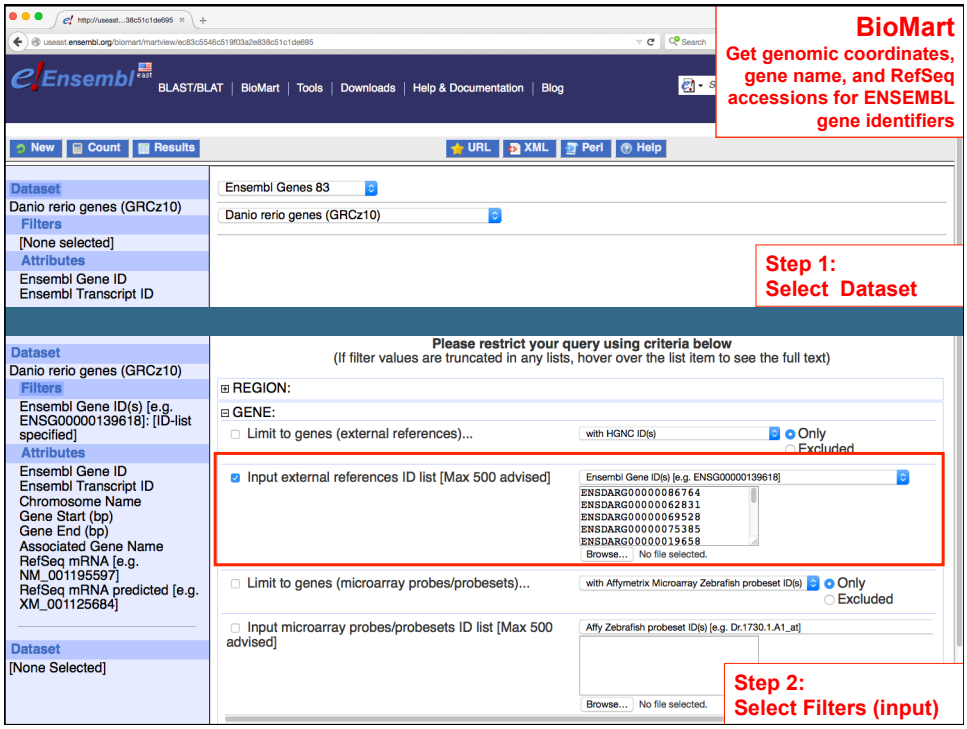
http://www.ensembl.org

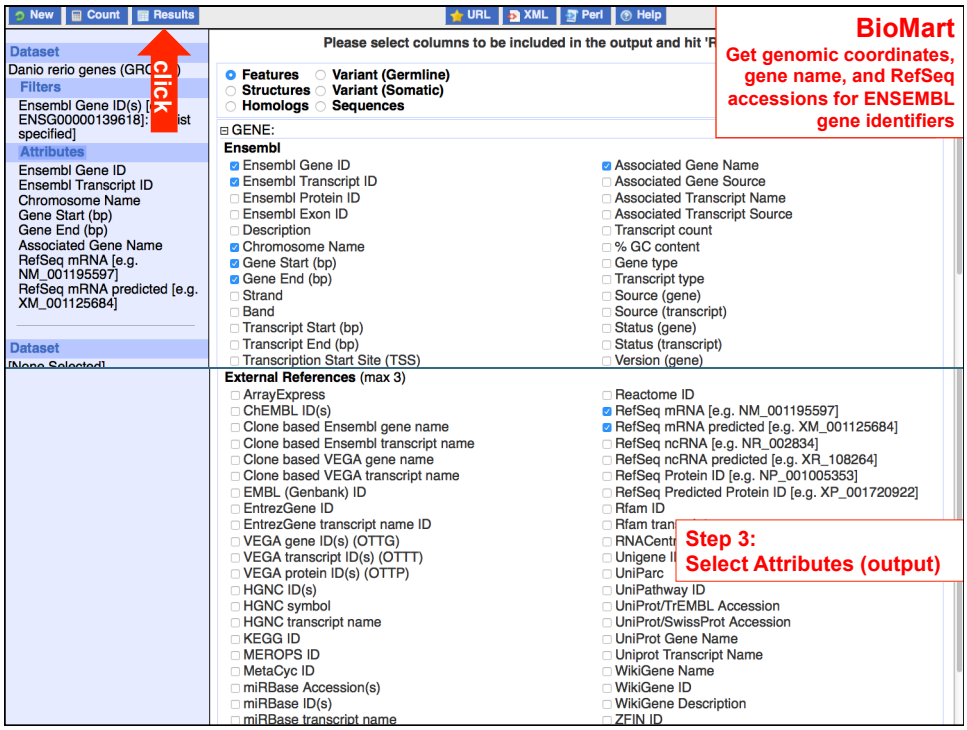
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

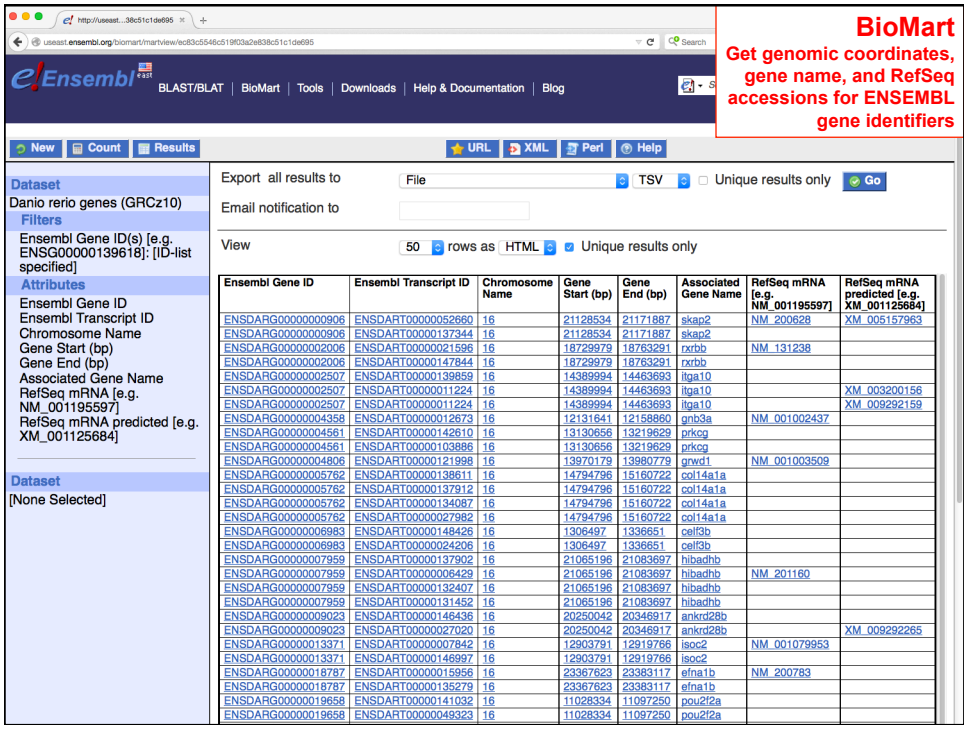



Ensembl BLAST results
Results table
Show All entries
Show/hide columns (2 hidden)
Filter
Genomic Location
Overlapping Gene(s)
Orientation
Query start
Query end
Length
Score
E-val
%ID
15:6235794-6237707 [Sequence] ENSGALG00000028177 Forward 26 662 656 [Sequence] 289 5e-87 34.45 [Alignment]
15:6239085-6241091 [Sequence] ENSGALG00000021341 Forward 2 672 690 [Sequence] 286 4e-86 33.62 [Alignment]
5:26725561-26727474 [Sequence] ENSGALG00000013838 Forward 27 663 654 [Sequence] 273 2e-81 31.80 [Alignment]
22:2444846-2446051 [Sequence] ENSGALG00000003444 Forward 183 406 405 [Sequence] 94.0 4e-19 24.44 [Alignment]
22:2457319-2458905 [Sequence] ENSGALG00000026763 Forward 339 647 535 [Sequence] 89.0 2e-27 23.55 [Alignment]
22:2452520-2453296 [Sequence] ENSGALG00000026763 Forward 438 603 260 [Sequence] 76.2 2e-13 28.08 [Alignment]
22:2451170-2452129 [Sequence] ENSGALG00000026763 Forward 219 403 325 [Sequence] 75.9 2e-13 23.69 [Alignment]
22:2450127-2450432 [Sequence] ENSGALG00000026763 Forward 92 169 102 [Sequence] 58.7 1e-07 39.22 [Alignment]
22:2446576-2446833 [Sequence] ENSGALG00000003444 Forward 438 534 98 [Sequence] 55.7 1e-06 38.78 [Alignment]
22:2456600-2456767 [Sequence] ENSGALG00000026763 Forward 215 270 56 [Sequence] 54.5 3e-09 53.57 [Alignment]
22:2435972-2436154 [Sequence] ADAM9 Forward 445 505 61 [Sequence] 53.6 5e-06 52.46 [Alignment]
1:195263819-195264010 [Sequence] ENSGALG00000017301 Forward
22:2456582-2457250 [Sequence] ENSGALG00000026763 Forward
22:2455569-2455859 [Sequence] ENSGALG00000026763 Forward
22:2435069-2435515 [Sequence] ADAM9 Forward
6:30944661-30944852 [Sequence] ENSGALG00000009606 Forward
6:32125158-32125331 [Sequence] ADAM12 Reverse
22:1049009-1049182 [Sequence] ENSGALG00000000357 Reverse
4:89097967-89098143 [Sequence] ADAM33 Reverse
13:10662489-10662740 [Sequence] ADAM19 Forward
6:8870586-8870750 [Sequence] ADAM8 Reverse
22:1049889-1050071 [Sequence] ENSGALG00000000357 Reverse
22:2447302-2447487 [Sequence] ENSGALG00000003444 Forward
6:30943023-30943193 [Sequence] ENSGALG00000009606 Forward
6:8869222-8869395 [Sequence] ADAM8 Reverse
22:2444167-2444289 [Sequence] ENSGALG00000003444 Forward
22:2456842-2457102 [Sequence] ENSGALG00000026763 Forward
27:1260914-1261267 [Sequence] ADAM11 Forward
22:2444028-2444099 [Sequence] ENSGALG00000003444 Forward
27:1260744-1260866 [Sequence] ADAM11 Forward
22:2435628-2435705 [Sequence] ADAM9 Forward
Query 26 QITVPEKIRSIIKEGIESQASYKIVIEGKPYTVNLM-QKNFLPHNFRVYSYSGTGIMKPL 84
Subject 6235794 EIVTPKKAGSKAGRASQGSMSYFISIQGVNYTIHLRHKKGFVVKNFPILTRDSEGQVMIE 6235973
Query 85 DQDFQNFCHYQGYIEGYPKSVVMVSTCTGLRGVLQFENVSYGIEPLESSVGFEHVIYQVK 144
Subject 6235974 QPRVLADCYYHGYVEGILDSTVTLTTCSGLRGLLQIGNLSYSIEPLAASSTFEHLLLQRE 6236153
Query 145 HKKADVSLYNEKDIESRDLSFKLQSVEPQQ----DFAKYIEMHVIVEKQLYNHMGSDTTV 200
Subject 6236154 AVVPGTVIYKTLQGGRR---FPGRGTAPRQFQPWGRTRYLELMVVVDKEGFDTFGTSITN 6236324
Query 201 VAQKVFQLIGLTNAIFVSFNITIILSSLELWIDENKIATTGEANELLHTFLRWKTSYLVL 260
Subject 6236325 VTLEVIEIINLVDGLFSSVRLRVLLTVLEIWTEKNPISITKNITQVLHSFNRWRIQHGPA 6236504
Query 261 R-PHDVAFLLVYREKSNYVGATFQG------KMCDANYAGGVVLHPRTISLESLAVILAQ 313
Subject 6236505 HIMHDVGCLFASLDFSRSTRALHVGGESNFASACNRQHSSAVVSFAKHTYIET-AVHVAH 6236681
Query 314 LLSLSMGITYDDINKCQCSGAV-CIMNPEAIHFSGVKIFSNCSFEDFAHFISKQKSQCLH 372
Subject 6236682 ELGYVLGMEHDD-EHCRCGNASKCIMNPKSTVSYG---FSNCSTKYYFDFITSGQGQCLN 6236849
Query 373 NQPR-LDPFFKQQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCRFKAGSNCAEGP 431
Subject 6236850 NIPSSIARFVPQR--CGNGVLEDREECDCGTEEQCK--SDPCCD-NTCRKKEGAICTSGQ 6237014
Query 432 CCENCLFMSKERMCRPSFEECDLPEYCNGSSASCPENHYVQTGHPCGLNQWICIDGVCMS 491
Subject 6237015 CCKDCKPLPEGVVCRKSTNPCDLPEYCNGTSEHCPEDVAKQDGTRCAADGY-CYSGKCRS 6237191

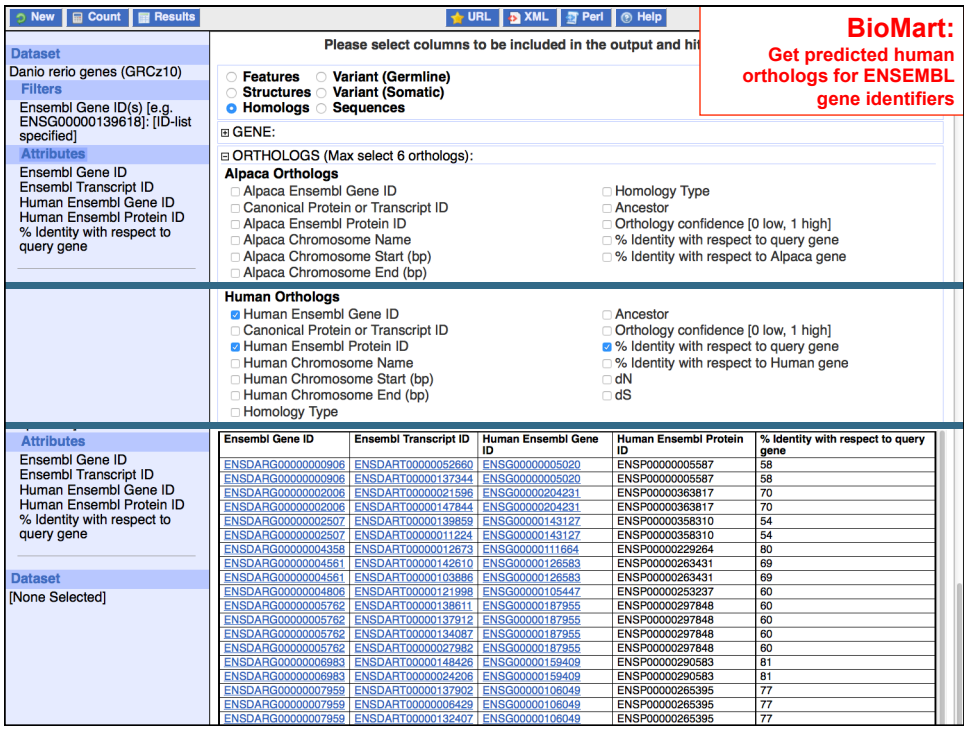

BioMart
Cross-reference data from different sources
http://www.ensembl.org/biomart
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research


BioMart
Get genomic coordinates, gene name, and RefSeq accessions for ENSEMBL gene identifiers
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog
New | Count | Results | URL | XML | Perl | Help
Dataset
Danio rerio genes (GRCz10)
Filters
[None selected]
Attributes
Ensembl Gene ID
Ensembl Transcript ID
Ensembl Genes 83
Danio rerio genes (GRCz10)
Step 1:
Select Dataset
Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)
REGION:
GENE:
Limit to genes (external references)...
Input external references ID list [Max 500 advised]
ENSDARG00000086764
ENSDARG00000062831
ENSDARG00000069528
ENSDARG00000075385
ENSDARG00000019658
Limit to genes (microarray probes/probesets)...
Input microarray probes/probesets ID list [Max 500 advised]
Step 2:
Select Filters (input)


BioMart
Get genomic coordinates, gene name, and RefSeq accessions for ENSEMBL gene identifiers
click
Please select columns to be included in the output and hit 'F
Features
Structures
Homologs
Variant (Germline)
Variant (Somatic)
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
Transcript Start (bp)
Transcript End (bp)
Transcription Start Site (TSS)
Associated Gene Name
Associated Gene Source
Associated Transcript Name
Associated Transcript Source
Transcript count
% GC content
Gene type
Transcript type
Source (gene)
Source (transcript)
Status (gene)
Status (transcript)
Version (gene)
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
EntrezGene transcript name ID
VEGA gene ID(s) (OTTG)
VEGA transcript ID(s) (OTTT)
VEGA protein ID(s) (OTTP)
HGNC ID(s)
HGNC symbol
HGNC transcript name
KEGG ID
MEROPS ID
MetaCyc ID
miRBase Accession(s)
miRBase ID(s)
Reactome 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
UniParc
UniPathway ID
UniProt/TrEMBL Accession
UniProt/SwissProt Accession
UniProt Gene Name
Uniprot Transcript Name
WikiGene Name
WikiGene ID
WikiGene Description
Step 3:
Select Attributes (output)



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	21128534	21171887	skap2	NM_200628	XM_005157963
ENSDARG00000000906	ENSDART00000137344	16	21128534	21171887	skap2		
ENSDARG00000002006	ENSDART00000021596	16	18729979	18763291	rxrbb	NM_131238	
ENSDARG00000002006	ENSDART00000147844	16	18729979	18763291	rxrbb		
ENSDARG00000002507	ENSDART00000139859	16	14389994	14463693	itga10		
ENSDARG00000002507	ENSDART00000011224	16	14389994	14463693	itga10		XM_003200156
ENSDARG00000002507	ENSDART00000011224	16	14389994	14463693	itga10		XM_009292159
ENSDARG00000004358	ENSDART00000012673	16	12131641	12158860	gnb3a	NM_001002437	
ENSDARG00000004561	ENSDART00000142610	16	13130656	13219629	prkcg		
ENSDARG00000004561	ENSDART00000103886	16	13130656	13219629	prkcg		
ENSDARG00000004806	ENSDART00000121998	16	13970179	13980779	grwd1	NM_001003509	
ENSDARG00000005762	ENSDART00000138611	16	14794796	15160722	col14a1a		
ENSDARG00000005762	ENSDART00000137912	16	14794796	15160722	col14a1a		
ENSDARG00000005762	ENSDART00000134087	16	14794796	15160722	col14a1a		
ENSDARG00000005762	ENSDART00000027982	16	14794796	15160722	col14a1a		
ENSDARG00000006983	ENSDART00000148426	16	1306497	1336651	celf3b		
ENSDARG00000006983	ENSDART00000024206	16	1306497	1336651	celf3b		
ENSDARG00000007959	ENSDART00000137902	16	21065196	21083697	hibadhb		
ENSDARG00000007959	ENSDART00000006429	16	21065196	21083697	hibadhb	NM_201160	
ENSDARG00000007959	ENSDART00000132407	16	21065196	21083697	hibadhb		
ENSDARG00000007959	ENSDART00000131452	16	21065196	21083697	hibadhb		
ENSDARG00000009023	ENSDART00000146436	16	20250042	20346917	ankrd28b		
ENSDARG00000009023	ENSDART00000027020	16	20250042	20346917	ankrd28b		XM_009292265
ENSDARG00000013371	ENSDART00000007842	16	12903791	12919766	isoc2	NM_001079953	
ENSDARG00000013371	ENSDART00000146997	16	12903791	12919766	isoc2		
ENSDARG00000018787	ENSDART00000015956	16	23367623	23383117	efna1b	NM_200783	
ENSDARG00000018787	ENSDART00000135279	16	23367623	23383117	efna1b		
ENSDARG00000019658	ENSDART00000141032	16	11028334	11097250	pou2f2a		
ENSDARG00000019658	ENSDART00000049323	16	11028334	11097250	pou2f2a		



Ensembl Gene ID	Ensembl Transcript ID	Human Ensembl Gene ID	Human Ensembl Protein ID	% Identity with respect to query gene
ENSDARG00000000906	ENSDART00000052660	ENSG00000005020	ENSP00000005587	58
ENSDARG00000000906	ENSDART00000137344	ENSG00000005020	ENSP00000005587	58
ENSDARG00000002006	ENSDART00000021596	ENSG00000204231	ENSP00000363817	70
ENSDARG00000002006	ENSDART00000147844	ENSG00000204231	ENSP00000363817	70
ENSDARG00000002507	ENSDART00000139859	ENSG00000143127	ENSP00000358310	54
ENSDARG00000002507	ENSDART00000011224	ENSG00000143127	ENSP00000358310	54
ENSDARG00000004358	ENSDART00000012673	ENSG00000111664	ENSP00000229264	80
ENSDARG00000004561	ENSDART00000142610	ENSG00000126583	ENSP00000263431	69
ENSDARG00000004561	ENSDART00000103886	ENSG00000126583	ENSP00000263431	69
ENSDARG00000004806	ENSDART00000121998	ENSG00000105447	ENSP00000253237	60
ENSDARG00000005762	ENSDART00000138611	ENSG00000187955	ENSP00000297848	60
ENSDARG00000005762	ENSDART00000137912	ENSG00000187955	ENSP00000297848	60
ENSDARG00000005762	ENSDART00000134087	ENSG00000187955	ENSP00000297848	60
ENSDARG00000005762	ENSDART00000027982	ENSG00000187955	ENSP00000297848	60
ENSDARG00000006983	ENSDART00000148426	ENSG00000159409	ENSP00000290583	81
ENSDARG00000006983	ENSDART00000024206	ENSG00000159409	ENSP00000290583	81
ENSDARG00000007959	ENSDART00000137902	ENSG00000106049	ENSP00000265395	77
ENSDARG00000007959	ENSDART00000006429	ENSG00000106049	ENSP00000265395	77
ENSDARG00000007959	ENSDART00000132407	ENSG00000106049	ENSP00000265395	77

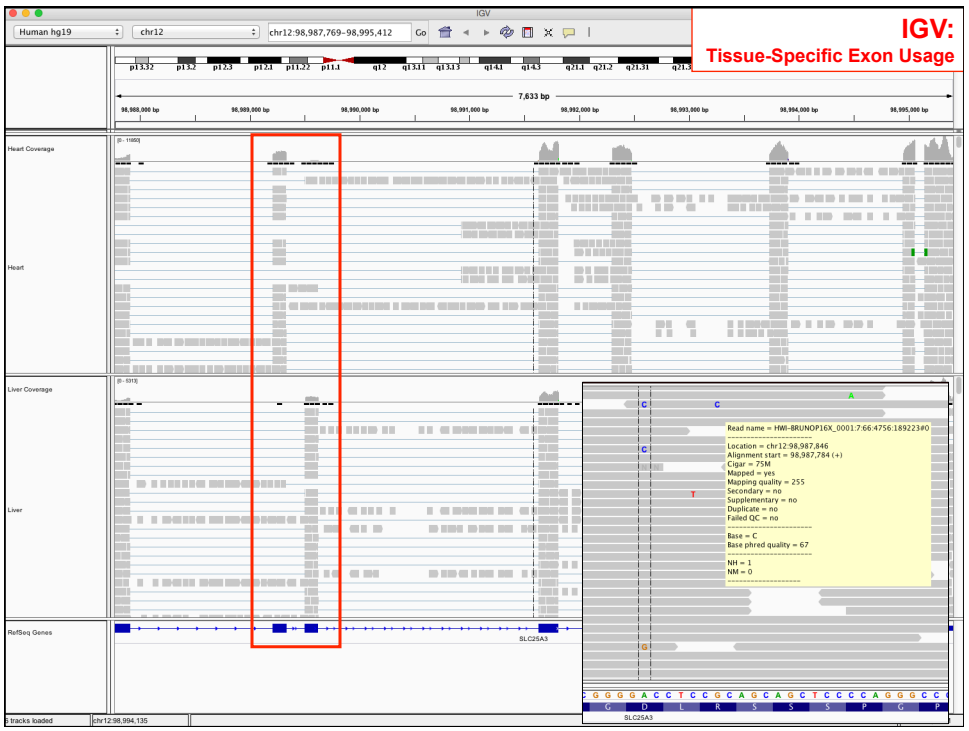
IGV

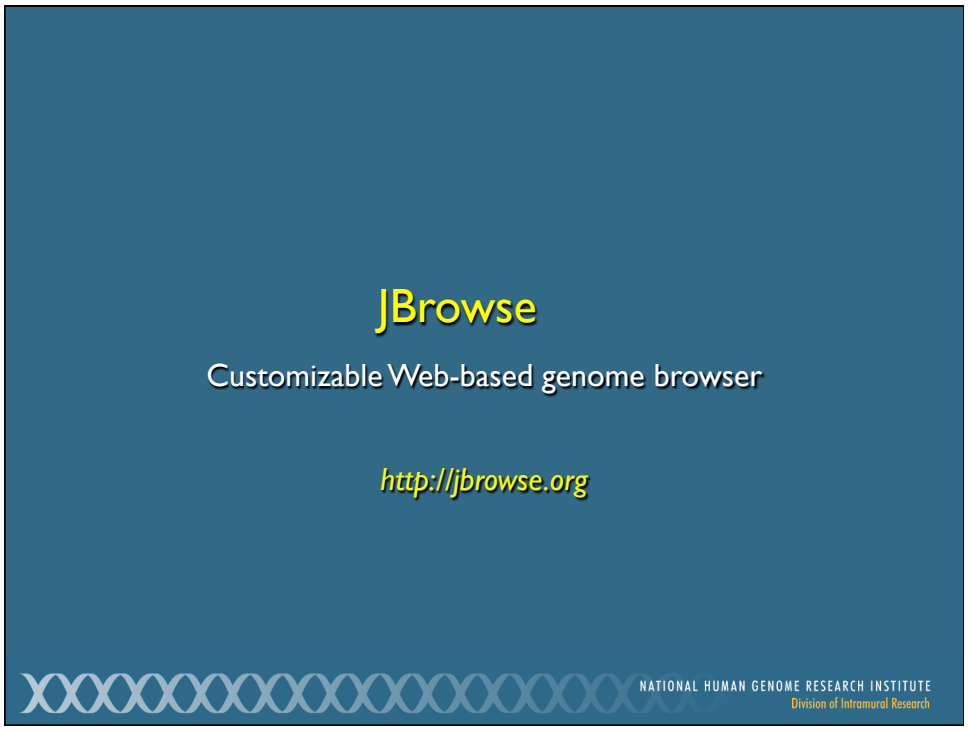
Customizable desktop genome browser

https://www.broadinstitute.org/igv

IGV

- Integrates sequence and array-based data, as well as clinical and phenotypic data
- Handles remote data from public servers, as well as local data generated by researchers
- Runs on a desktop, and designed for both bench biologists and bioinformaticians

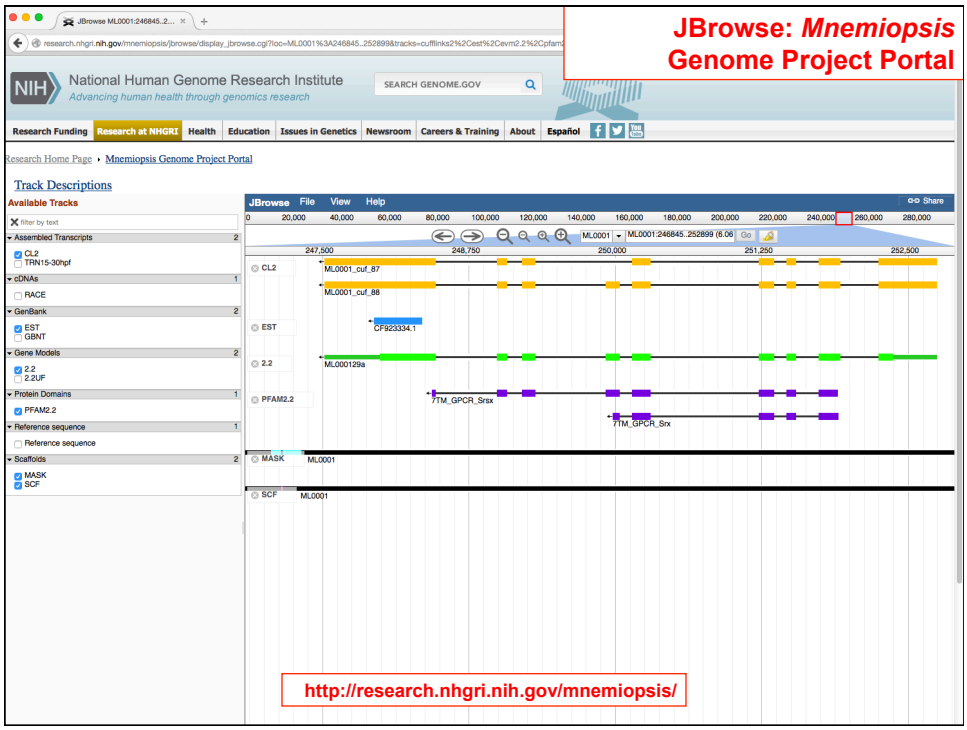

IGV:
Tissue-Specific Exon Usage
IGV
Human hg19
chr12
chr12:98,987,769-98,995,412
Go
7,633 bp
Heart Coverage
Heart
Liver Coverage
Liver
RefSeq Genes
SLC25A3
Read name = HWI-BRUNOP16X_0001:7:66:4756:189223#0
Location = chr12:98,987,846
Alignment start = 98,987,784 (+)
Cigar = 75M
Mapped = yes
Mapping quality = 255
Secondary = no
Supplementary = no
Duplicate = no
Failed QC = no
Base = C
Base phred quality = 67
NH = 1
NM = 0


JBrowse
Customizable Web-based genome browser
http://jbrowse.org
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

JBrowse

- Customizable Web-based genome browser for visualizing large-scale genomic data
- Can be configured to display any genome and associated data and annotations
- Requires bioinformatics expertise and hardware, especially to share data

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

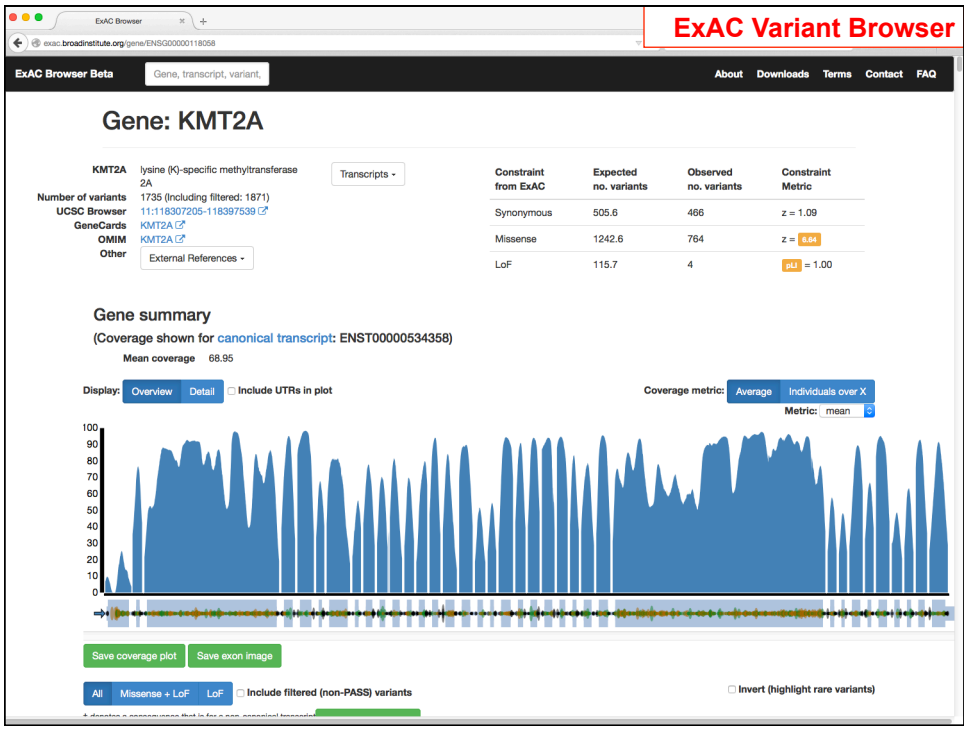


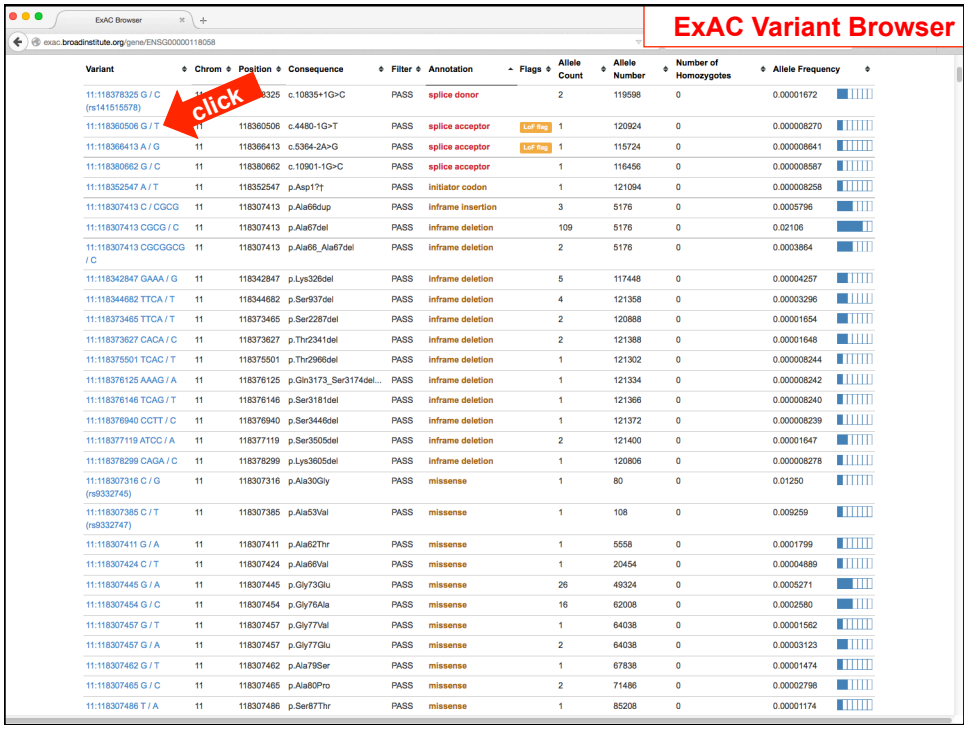
Exome Aggregation Consortium (ExAC)

Exome sequencing data from 60,000 individuals

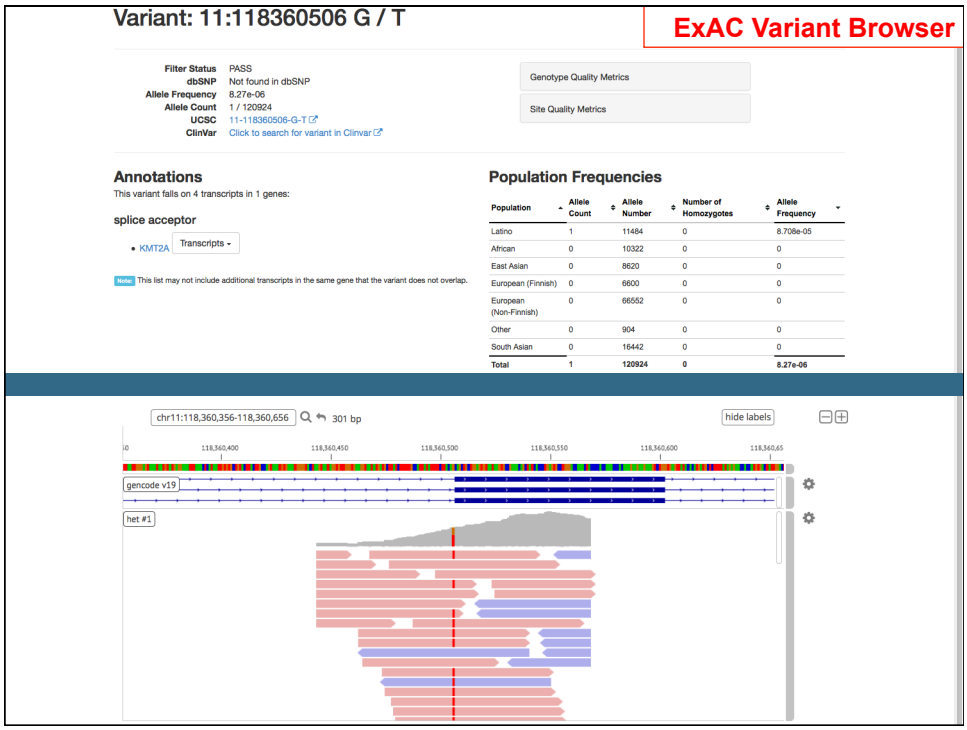
http://exac.broadinstitute.org

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research





Variant	Chrom	Position	Consequence	Filter	Annotation	Flags	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
11:118378325 G / C (rs141515578)	11	118378325	c.10835+1G>C	PASS	splice donor		2	119598	0	0.00001672
11:118360506 G / T	11	118360506	c.4480-1G>T	PASS	splice acceptor	LoF flag	1	120924	0	0.000008270
11:118366413 A / G	11	118366413	c.5364-2A>G	PASS	splice acceptor	LoF flag	1	115724	0	0.000008641
11:118380662 G / C	11	118380662	c.10901-1G>C	PASS	splice acceptor		1	116456	0	0.000008587
11:118352547 A / T	11	118352547	p.Asp1?†	PASS	initiator codon		1	121094	0	0.000008258
11:118307413 C / CGCG	11	118307413	p.Ala66dup	PASS	inframe insertion		3	5176	0	0.0005796
11:118307413 CGCG / C	11	118307413	p.Ala67del	PASS	inframe deletion		109	5176	0	0.02106
11:118307413 CGCGGCG / C	11	118307413	p.Ala66_Ala67del	PASS	inframe deletion		2	5176	0	0.0003864
11:118342847 GAAA / G	11	118342847	p.Lys326del	PASS	inframe deletion		5	117448	0	0.00004257
11:118344682 TTCA / T	11	118344682	p.Ser937del	PASS	inframe deletion		4	121358	0	0.00003296
11:118373465 TTCA / T	11	118373465	p.Ser2287del	PASS	inframe deletion		2	120888	0	0.00001654
11:118373627 CACA / C	11	118373627	p.Thr2341del	PASS	inframe deletion		2	121388	0	0.00001648
11:118375501 TCAC / T	11	118375501	p.Thr2966del	PASS	inframe deletion		1	121302	0	0.000008244
11:118376125 AAAG / A	11	118376125	p.Gln3173_Ser3174del...	PASS	inframe deletion		1	121334	0	0.000008242
11:118376146 TCAG / T	11	118376146	p.Ser3181del	PASS	inframe deletion		1	121366	0	0.000008240
11:118376940 CCTT / C	11	118376940	p.Ser3446del	PASS	inframe deletion		1	121372	0	0.000008239
11:118377119 ATCC / A	11	118377119	p.Ser3505del	PASS	inframe deletion		2	121400	0	0.00001647
11:118378299 CAGA / C	11	118378299	p.Lys3605del	PASS	inframe deletion		1	120806	0	0.000008278
11:118307316 C / G (rs9332745)	11	118307316	p.Ala30Gly	PASS	missense		1	80	0	0.01250
11:118307385 C / T (rs9332747)	11	118307385	p.Ala53Val	PASS	missense		1	108	0	0.009259
11:118307411 G / A	11	118307411	p.Ala62Thr	PASS	missense		1	5558	0	0.0001799
11:118307424 C / T	11	118307424	p.Ala66Val	PASS	missense		1	20454	0	0.00004889
11:118307445 G / A	11	118307445	p.Gly73Glu	PASS	missense		26	49324	0	0.0005271
11:118307454 G / C	11	118307454	p.Gly76Ala	PASS	missense		16	62008	0	0.0002580
11:118307457 G / T	11	118307457	p.Gly77Val	PASS	missense		1	64038	0	0.00001562
11:118307457 G / A	11	118307457	p.Gly77Glu	PASS	missense		2	64038	0	0.00003123
11:118307462 G / T	11	118307462	p.Ala79Ser	PASS	missense		1	67838	0	0.00001474
11:118307465 G / C	11	118307465	p.Ala80Pro	PASS	missense		2	71486	0	0.00002798
11:118307486 T / A	11	118307486	p.Ser87Thr	PASS	missense		1	85208	0	0.00001174

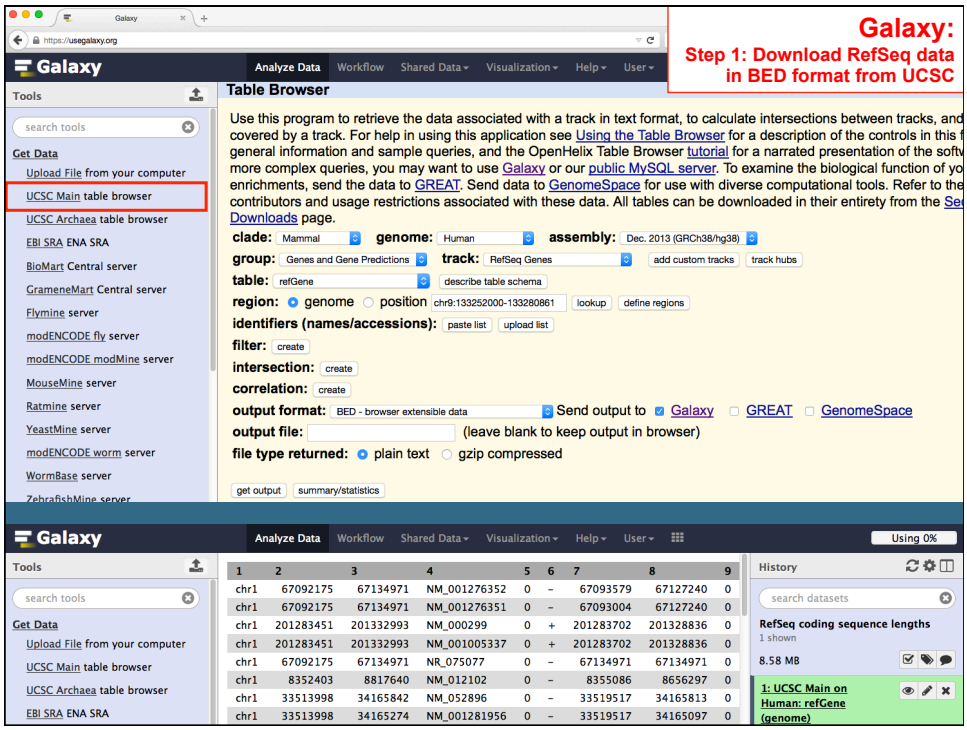


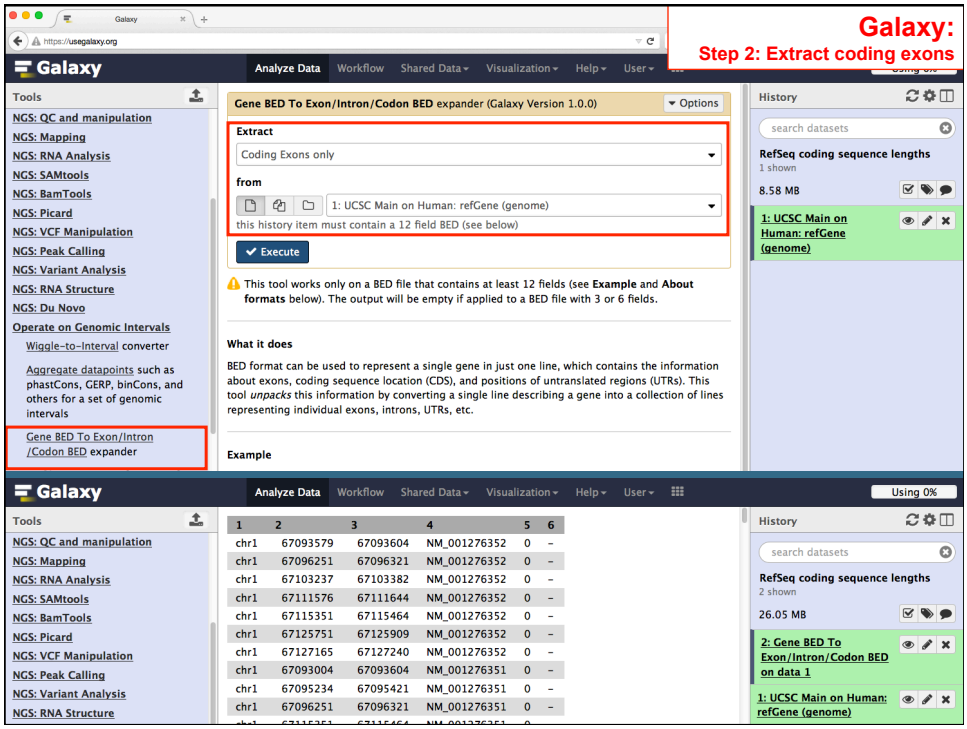
Galaxy

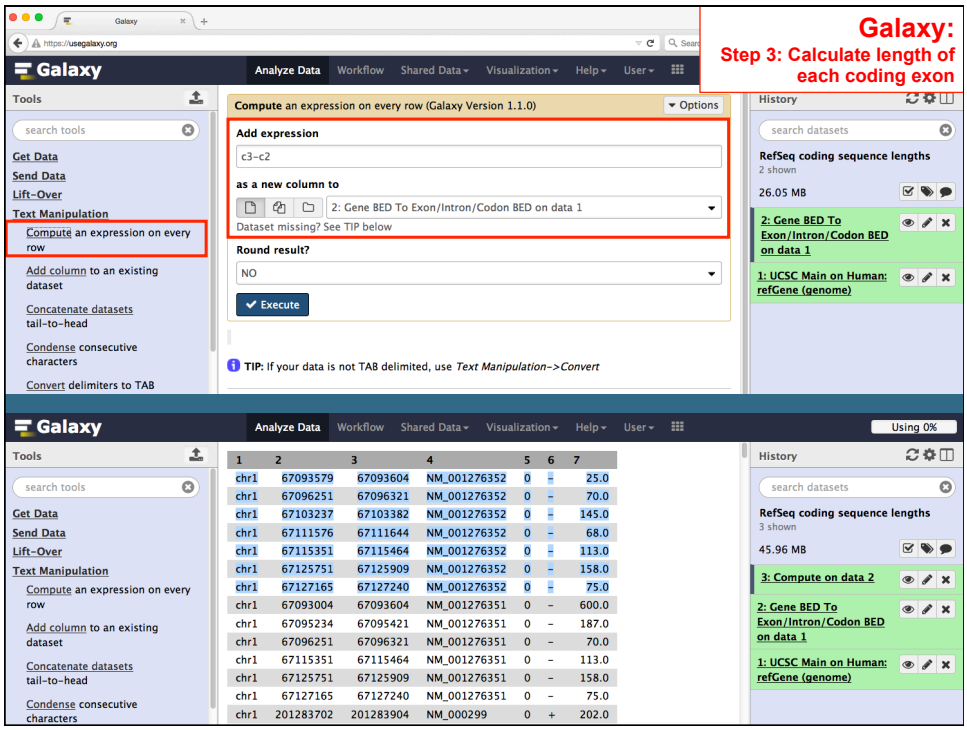
Web-based genome analysis

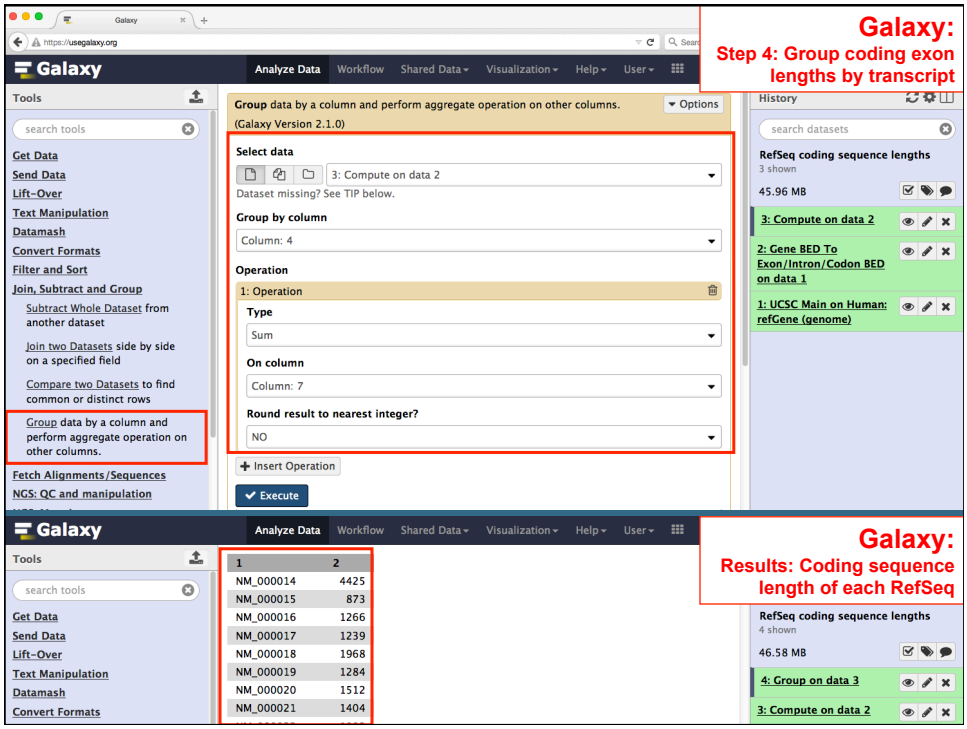
https://usegalaxy.org

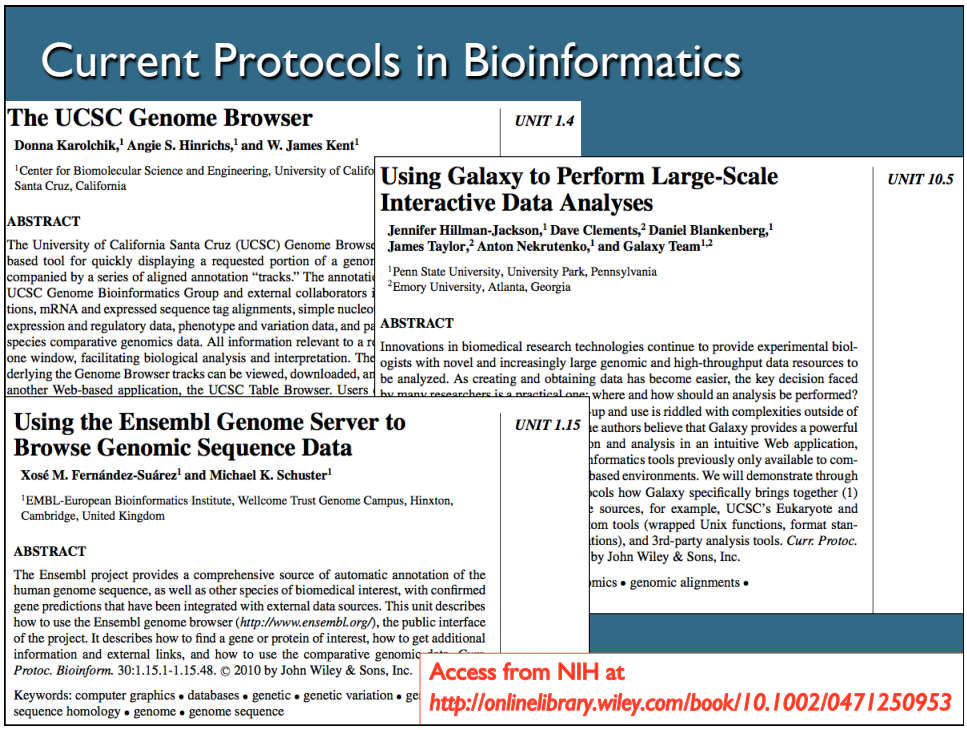
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research




Galaxy:
Step 2: Extract coding exons
Gene BED To Exon/Intron/Codon BED expander (Galaxy Version 1.0.0)
Extract
Coding Exons only
from
1: UCSC Main on Human: refGene (genome)
this history item must contain a 12 field BED (see below)
Execute
What it does
Example
History
RefSeq coding sequence lengths
1 shown
8.58 MB
1: UCSC Main on Human: refGene (genome)
2 shown
26.05 MB
2: Gene BED To Exon/Intron/Codon BED on data 1


Galaxy:
Step 3: Calculate length of each coding exon
Compute an expression on every row (Galaxy Version 1.1.0)
Add expression
c3-c2
as a new column to
2: Gene BED To Exon/Intron/Codon BED on data 1
Round result?
NO
Execute
TIP: If your data is not TAB delimited, use Text Manipulation->Convert
3 shown
45.96 MB
3: Compute on data 2


Galaxy:
Step 4: Group coding exon lengths by transcript
Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Tools
search tools
Get Data
Send Data
Lift-Over
Text Manipulation
Datamash
Convert Formats
Filter and Sort
Join, Subtract and Group
Subtract Whole Dataset from another dataset
Join two Datasets side by side on a specified field
Compare two Datasets to find common or distinct rows
Group data by a column and perform aggregate operation on other columns.
Fetch Alignments/Sequences
NGS: QC and manipulation
Group data by a column and perform aggregate operation on other columns.
(Galaxy Version 2.1.0)
Options
Select data
3: Compute on data 2
Dataset missing? See TIP below.
Group by column
Column: 4
Operation
1: Operation
Type
Sum
On column
Column: 7
Round result to nearest integer?
NO
Insert Operation
Execute
History
search datasets
RefSeq coding sequence lengths
3 shown
45.96 MB
3: Compute on data 2
2: Gene BED To Exon/Intron/Codon BED on data 1
1: UCSC Main on Human: refGene (genome)
Galaxy:
Results: Coding sequence length of each RefSeq
1 2
NM_000014 4425
NM_000015 873
NM_000016 1266
NM_000017 1239
NM_000018 1968
NM_000019 1284
NM_000020 1512
NM_000021 1404
RefSeq coding sequence lengths
4 shown
46.58 MB
4: Group on data 3
3: Compute on data 2


Current Protocols in Bioinformatics
The UCSC Genome Browser
UNIT 1.4
Donna Karolchik, Angie S. Hinrichs, and W. James Kent
Center for Biomolecular Science and Engineering, University of California, Santa Cruz, California
ABSTRACT
Using Galaxy to Perform Large-Scale Interactive Data Analyses
UNIT 10.5
Jennifer Hillman-Jackson, Dave Clements, Daniel Blankenberg, James Taylor, Anton Nekrutenko, and Galaxy Team
Penn State University, University Park, Pennsylvania
Emory University, Atlanta, Georgia
ABSTRACT
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
Using the Ensembl Genome Server to Browse Genomic Sequence Data
UNIT 1.15
Xosé M. Fernández-Suárez and Michael K. Schuster
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