

Worked example: From a chromosomal region to a gene

Ensembl Genome Browser

About Ensembl!

Ensembl is a joint project between [EBI](#), [Sanger Institute](#) and the [Wellcome Trust](#) to develop a software system which produces and maintains automatic annotation on eukaryotic genomes. Ensembl is primarily funded by the [Wellcome Trust](#). Access to all the data produced by the project, and to the software used to analyse and present it, is provided free and without constraints.

Ensembl presents up-to-date sequence data and the best possible automatic annotation for eukaryotic genomes. Available now are [human](#), [mouse](#) and [Drosophila](#). Others will be added soon.

For an introduction to the Ensembl project, take the [Ensembl tour](#). For more information read this short [paper](#) in Nucleic Acids Research.

For all enquiries, please contact the Ensembl [Help Desk](#) (helpdesk@ensembl.org).

Ensembl provides

- Easy access to sequence data
- For known genes, predicted structure and location in the genome sequence
- Prediction of novel genes, all with supporting evidence
- Annotation of other features of the genome
- Targetted connections to other genome resources worldwide

Easy access to the data via

- A web-based genome browser (which can be customized as required)
- A web-based system for data export and data mining
- 'Dumps' of sequence and other data sets for you to download
- Direct access to the databases
- A Perl-based object layer

Ensembl Species

Species	Version	Date
Human	v. 4.28.1	12 Mar 2002
Mouse	v. 4.1.1	31 Jan 2002
Fly	v. 4.1.1	11 Feb 2002

Access to whole genome shotgun data (includes additional species) [Trace Server](#)

Help and documentation

- Species-specific documentation is available via the species home pages above.
- Take the [Ensembl tour](#) or read this short [paper](#) in Nucleic Acids Research.
- For context-sensitive help on any web page click: [Help](#)
- There is also an [index](#) of context-sensitive help pages, and a set of guided [How do I...?](#) trails.

Recent Ensembl news

Display your own data in Ensembl [News](#)

Ensembl developer resources [DAS](#)

Apollo genome browser [Development](#)

Questions or suggestions? Try the [Apollo](#)

Documentation (includes tutorial on direct data access & instructions for installing Ensembl on your own site) [Help Desk](#)

[Documentation](#)

Have you tried ...?

Data Export - Gene List

Export list of gene data in text, html or Excel (tm) format

[Click for more information](#)

Pick Human Ensembl

'About Ensembl' plus general help and documentation – multi-species

Human Genome Browser

Ensembl Entry Points

Search for with [Lookup](#)

Display Chr From To [Lookup](#)

Retrieve a sequence [Export Sequence](#)

BLAST your sequence [Blast](#)

Request a gene or SNP list [Export Data](#)

For fast identity search try [SSAHA](#)

Browse a Chromosome

Documentation & Help

About Ensembl [Home](#)

For context-sensitive help on any web page click [Help](#)

Questions or suggestions? Try [Help Desk](#)

Ensembl Links and Site Map

[Download](#)

[Export Sequence](#)

[Export Data](#)

[Blast](#)

[SSAHA](#)

[Site Map](#)

Other Species

[Mouse](#) [Fly](#)

Current Release 4.28.1

This release is based on the NCBI 28 assembly of the human genome.

Last Update: 07-03-2002

Ensembl gene predictions: 24179

Ensembl gene predictions: 27720

Ensembl gene exons: 199740

Ensembl gene transcripts: 29076

Contigs: 201753

Clones: 30050

Base Pairs: 4135116206

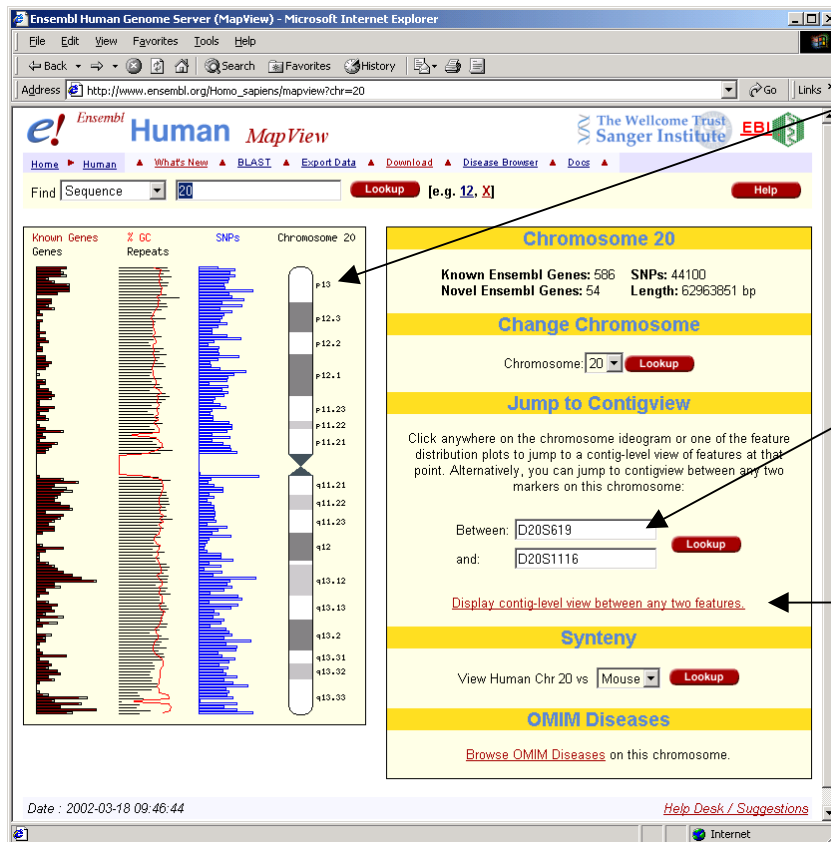
[What's New](#)

Search function

STEP 1
Pick chromosome 20

Help!

Quick access to the different "Views"

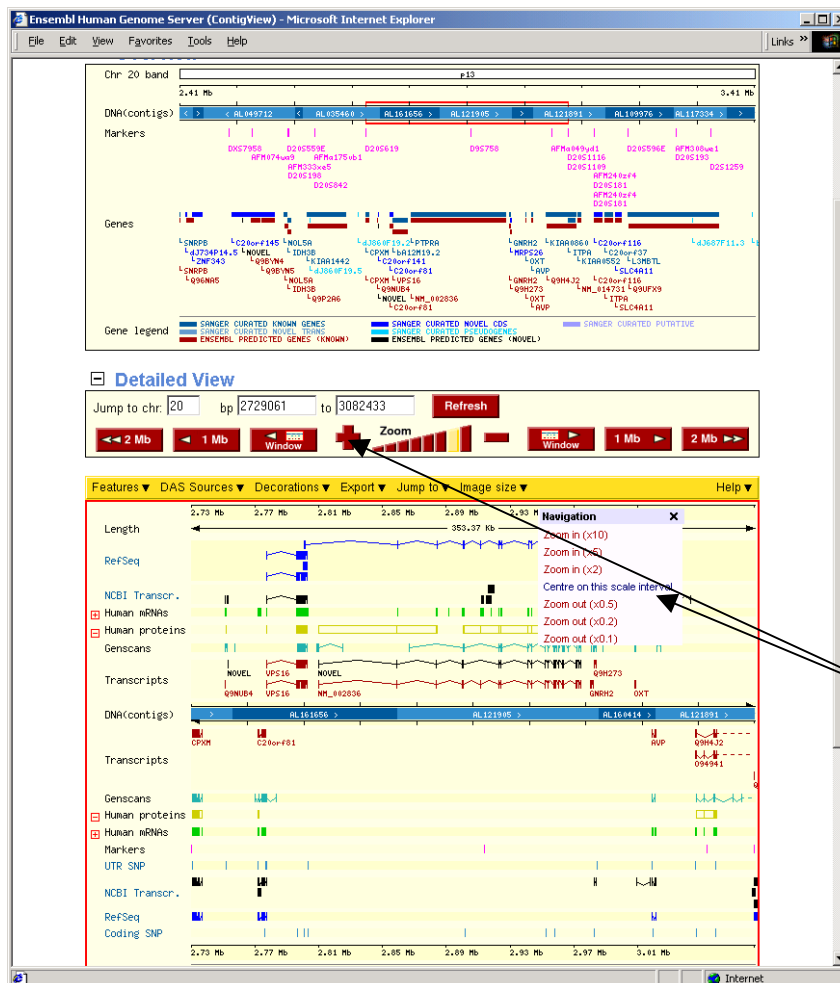


Display any region of interest

STEP 2

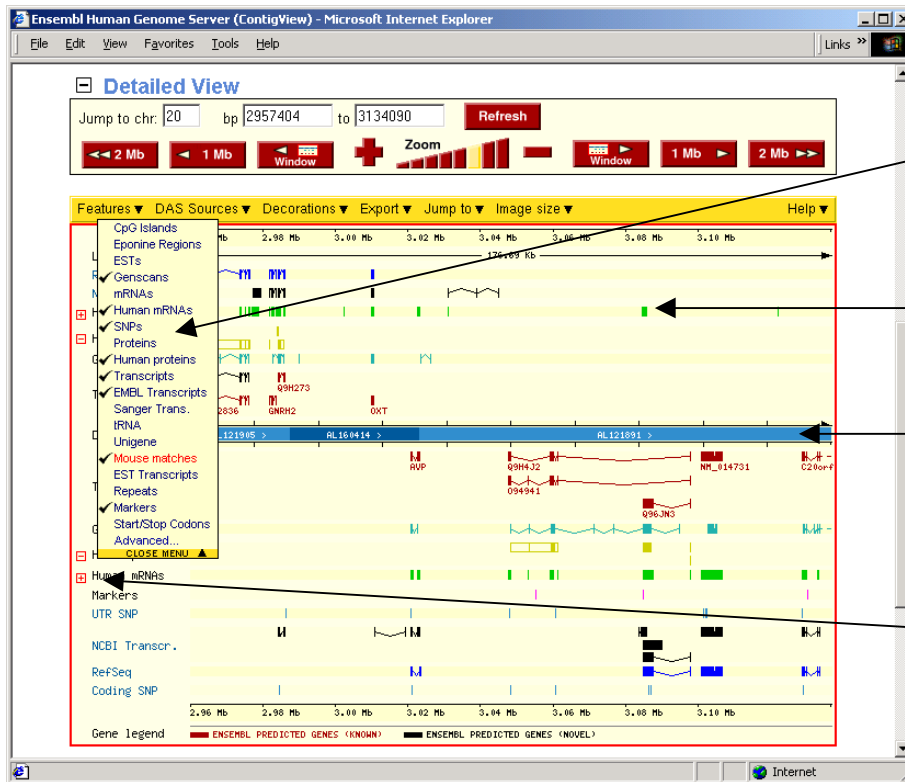
Display the region between markers D20S619 and D20S1116

Display a region between any two features



STEP 3

Centre around the AVP gene and then zoom in using the "plus" button



STEP 4

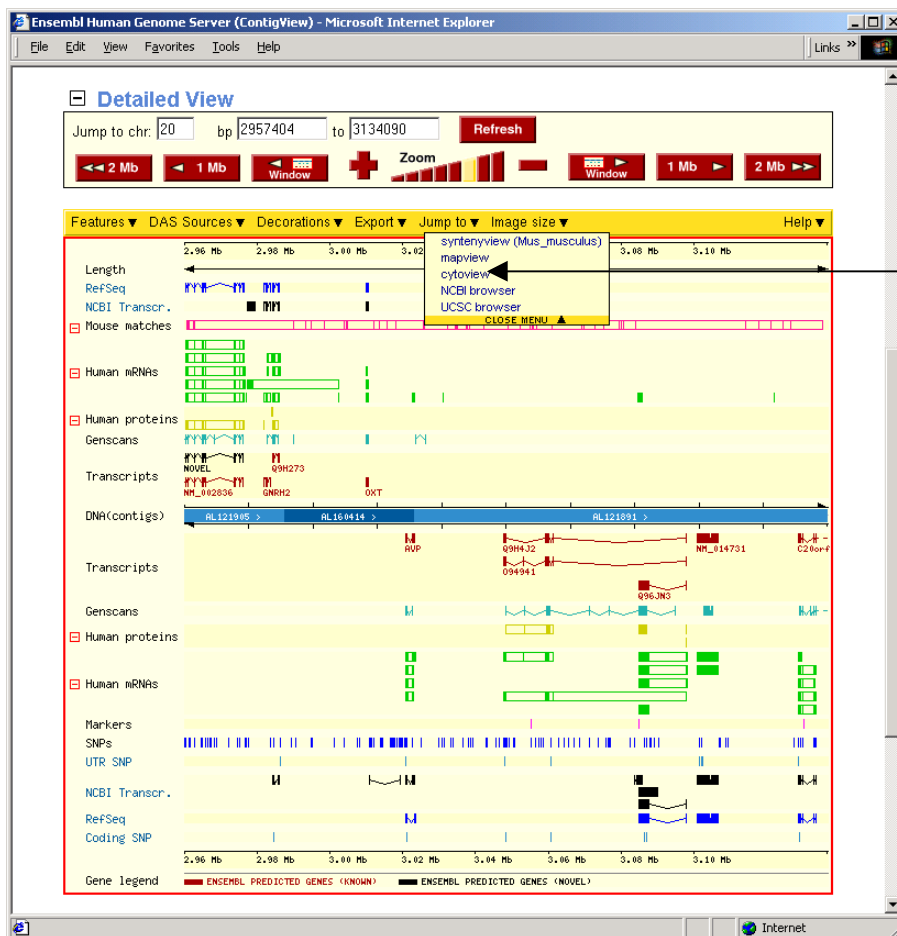
Customise the display – show SNPs and mouse matches

Evidence for gene prediction

DNA sequence - contigs

STEP 5

‘Expand’ the Human mRNA evidence



STEP 6

Jump to “cytoview”

Ensembl Human Genome Server (CytoView) - Microsoft Internet Explorer

Home Human What's New BLAST Export Data Download Disease Browser Docs

Find Sequence Lookup [e.g. AP000869, RH9632, cancer] Help

Chromosome 20

Chr 20

Detailed View

Jump to chr: 20 bp 2957404 to 3134090 Refresh Band: Refresh

<< 2 Mb < 1 Mb Zoom + > 1 Mb > 2 Mb >>

show 2.00 Mb in detail

Features DAS Sources Decorations Export Jump to Image size Help

Length 176.69 Kb

Chr 20 band

DNA(contigs) AL121995 AL164414

1Mb cloneset

Tile Path

2.96 Mb 3.13 Mb

Genes

Repeats

Repeats(Alu)

Repeats(LINE)

Gene legend

As HTML CSV

Date: 2002-03-18 14:36:58 Help Desk / Suggestions

http://www.ensembl.org/Homo_sapiens/cytoview?chr=20&vc_start=2045747&vc_end=4045747

STEP 7

Zoom into larger view

Dump clones and genes

STEP 8

Go **BACK** to "ContigView" and then click on a marker

Ensembl Human Genome Server (MarkerView) - Microsoft Internet Explorer

Home Human What's New BLAST Export Data Download Disease Browser Docs

Find Marker Lookup [e.g. RH9632, D1S2806] Help

Chromosome Map Marker D20S1109

Assembly location: D20S1109 is currently mapped to the following Ensembl location(s):

Chromosomal position	Flat file of region
20 (3082316 bp - 3082433 bp)	Export data

Marker details:

Name	Database
D20S1109	GDB

Marker D20S1109 has 11 synonym(s):

Name	Database
23091	dbSTS
gdb:738443	GDB
gdb:737397	GDB
STS12947	RHdb
SU7320	RHalloc
SHGC:11829	SHGC
T62632	EMBL
RH8701	RHdb
RH95868	RHdb
DBSTS:23091	DBSTS
G14672	EMBL

Marker D20S1109 primers:

Left Primer	Right Primer
CCTCAACACACTATTTTATCGCC	TCTGAAGGGCATTAGGGG

Date: 2002-03-18 17:03:13 Help Desk / Suggestions

Done Internet

STEP 9

Go **BACK** and then click on the AVP transcript

Ensembl Human Genome Server (ProtView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help Links

Ensembl Protein Report

Ensembl Protein	ENSP00000217172
Ensembl Gene	This protein is a product of Ensembl gene ENSG00000101200 [Supporting evidence]
Description	NEUROPHYSIN PRECURSOR [CONTAINS: NEUROPHYSIN [Source:ENSEMBL_PROTEIN_FAMILIES;Acc:ENSF00000000371]]
Prediction Method	This protein was predicted by the Ensembl analysis pipeline from either a GeneWise or Genscan prediction followed by confirmation of the exons by comparisons to protein, cDNA and EST databases
InterPro	IPR000981 Neurohypophysial hormone [View other Ensembl genes with this domain]
Protein Family	ENSF00000000371 NEUROPHYSIN PRECURSOR [CONTAINS: NEUROPH This cluster contains 2 Ensembl gene member(s)
Protein structure	

Peptide : ENSP00000217172

Peptide sequence

```
>ENSP00000217172
HPUTMLPACFLGLLAFSSACYFONCFRGGKRAMSDLELRQCLPGFGGKGRCPGPSICCADELGCFWGT
EALPCQENYILFSPCCSGKAGSGGRCAAFGVCNDSCVTEPECREGFIRRRASDRSNATQLDGFAG
AIIILRLVQLAGAPEFFPEAQPDA
```

Peptide properties

Residues: 164
MW: 17324.67
Avg. Res. Wt.: 105.638
Charge: -2.5
pI: 4.9366

[View Transcript Info](#)

This peptide corresponds to the following identifiers with the percent identity specified:

RefSeq: [NM_000490](#) [Target % identity : 100; Query % identity : 100]

Domains	Prediction method	Accession number	Start	End	Description
prints		PR00831	21	35	Neurhypophys_horm
prints		PR00831	39	55	Neurhypophys_horm
prints		PR00831	55	74	Neurhypophys_horm

STEP 11

Follow link to protein family

Ensembl Human Genome Server (FamilyView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help Links

Ensembl Human FamilyView

Home Human What's New BLAST Export Data Download Disease Browser Post

Find Family [Lookup](#) [e.g. [ENSF00000000117](#)] [Help](#)

Ensembl Protein Family ENSF00000000371 members

Consensus Annotation: NEUROPHYSIN PRECURSOR [CONTAINS: NEUROPHYSIN.

The annotation confidence score of this family is 71.

Protein families generated by TRIBE-MCL: Enright A.J., Van Dongen S. and Ouzounis C.A. "An efficient algorithm for large-scale detection of protein families." [paper submitted 2002]

The following Ensembl genes contain peptides in family ENSF00000000371

Export a list of genes containing this family
Please click on the gene identifier to go to a graphical gene view

Chr.	Ensembl Gene	Description
20	ENSG00000101200	NEUROPHYSIN PRECURSOR [CONTAINS: NEUROPHYSIN [Source:ENSEMBL_PROTEIN_FAMILIES;Acc:ENSF00000000371]]
20	ENSG00000101405	OXYTOCIN-NEUROPHYSIN 1 PRECURSOR (OT-NP) [CONTAINS: OXYTOCIN/OXYTOCIN; NEUROPHYSIN 1]. [Source:SWISSPROT;Acc:P01178]

SWISSPROT & SPTREMBL					
Q14935	P01179	P08163	P16229	P81768	Q91226
O12493	P01180	P11769	D17692	O07667	O01501

Chromosomal location and links to other family members

STEP 12

Go BACK to "GeneView" and then follow link to LocusLink

LocusLink Report - Microsoft Internet Explorer

File Edit View Favorites Tools Help Links »

NCBI **LocusLink**

PubMed Entrez BLAST OMIM Taxonomy Structure

Search LocusLink Display Brief Organism: All

Query: Go Clear

View Hs AVP One of 1 Loci Save All Loci

ABCDEFGHIJKLMNOPQRSTUVWXYZ

Click to Display mRNA-Genomic Alignments (spanning 2872 bps)

PUB OMIM UNIGENE MAP VAR HOMOL GDB HGMD

e! UCSC PROTEOME

Homo sapiens Official Gene Symbol and Name (HGNC)

AVP: arginine vasopressin (neurophysin II, antidiuretic hormone, diabetes insipidus, neurohypophyseal)

LocusID: 551

Overview ?

RefSeq Summary: This gene encodes a precursor protein consisting of arginine vasopressin and two associated proteins, neurophysin II and a glycopeptide, copeptin. Arginine vasopressin is a posterior pituitary hormone which is synthesized in the supraoptic nucleus and paraventricular nucleus of the hypothalamus. Along with its carrier protein, neurophysin II, it is packaged into neurosecretory vesicles and transported axonally to the nerve endings in the neurohypophysis where it is either stored or secreted into the bloodstream. The precursor is thought to be activated while it is being transported along the axon to the posterior pituitary. Arginine vasopressin acts as a growth factor by enhancing pH regulation through acid-base transport systems. It has a direct antidiuretic action on the kidney, and also causes vasoconstriction of the peripheral vessels. This hormone can contract smooth muscle during parturition and lactation. It is also involved in cognition, tolerance, adaptation and complex sexual and maternal behaviour, as well as in the regulation of water excretion and cardiovascular functions. Mutations in this gene cause autosomal dominant neurohypophyseal diabetes insipidus (ADNDI).

Proteome Summary: Prepro-arginine-vasopressin-neurophysin II, prohormone

Locus Type: gene with protein product, function known or inferred

Product: arginine vasopressin-neurophysin II preproprotein

AVP-NP11

copeptin

(All Pubs) ?

Source	Pub
Proteome	pm
Proteome	pm
Proteome	pm
Proteome	pm
Proteome	pm
Proteome	pm
Proteome	pm

Source	Pub
Proteome	pm
Proteome	pm
Proteome	pm

?

STEP 13

Go **BACK** and then follow link to OMIM (MIM)

Mouse Homology Maps:

NCBI vs. MGD	2 73.20 cM	Avp	Hs Mm
UCSC vs. MGD	2 73.20 cM	Avp	Hs Mm

Map Information ?

Chromosome: 20 **mv**

Cytogenetic: 20p13 [HUGO](#)

NCBI Reference Sequences (RefSeq) ?

Category: REVIEWED

mRNA: [NM_000490](#)

Protein: [NP_000481](#) arginine vasopressin-neurophysin II **BL**
preproprotein

Domains: [Neurohypophysial hormones](#) score: 252
[Neurohypophysial hormones](#) score: 271

GenBank [AF031475_M11166](#)

Source:

Category: NCBI Genome Annotation

Genomic: [NT_011387](#) **sv mv ev**

OMIM ENTRY 192340 - Microsoft Internet Explorer

File Edit View Favorites Tools Help Links »

OMIM Home Search Comments

***192340 ARGININE VASOPRESSIN; AVP**

Alternative titles; symbols

ARVP
VASOPRESSIN-NEUROPHYSIN II
ANTIDIURETIC HORMONE; ADH
NEUROPHYSIN II, INCLUDED; NPII, INCLUDED

TABLE OF CONTENTS

- [TEXT](#)
- [ALLELIC VARIANTS](#)
 - [View List of allelic variants](#)
- [REFERENCES](#)
- [SEE ALSO](#)
- [CONTRIBUTORS](#)
- [CREATION DATE](#)
- [EDIT HISTORY](#)
- [MINI-MIM](#)
- [CLINICAL SYNOPSIS](#)

Database Links

MEDLINE Protein DNA HGMD LocusLink Gene Map GDB MGD Nonenclature

Gene Map Locus: [20p13](#)

Note: pressing the ⓘ symbol will find the citations in MEDLINE whose text most closely matches the text of the preceding OMIM paragraph, using the Entrez MEDLINE neighboring function.

TEXT

[Sachs et al. \(1969\)](#) suggested that arginine vasopressin and its corresponding neurophysin are synthesized in the form of a common precursor which is cleaved by proteolysis to yield the biologically functional peptides. Rats with hereditary diabetes insipidus are deficient in synthesis of both arginine vasopressin and one species of neurophysin ([Sunde and Sokol, 1975](#)). Both of the nonapeptide hormones arginine vasopressin and oxytocin (OXT; [167050](#)) are synthesized in the supraoptic nucleus (SON) and paraventricular nucleus (PVN) of the hypothalamus together with their respective 'carrier' proteins, the neurophysins ([Brownstein et al., 1980](#)). Vasopressin and oxytocin are produced by separate populations of magnocellular neurons in both nuclei. Together with the neurophysins they are packaged into neurosecretory vesicles and transported axonally to the nerve endings in the neurohypophysis, where they are either stored or secreted into the bloodstream. In addition to having 9 amino acids, each has a

STEP 14

Go **BACK** and then click on “Export gene data in EMBL, GenBank or FASTA”

Ensembl Human Genome Server (ExportView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help Links »

e! Ensembl Human ExportView

The Wellcome Trust Sanger Institute EBI

Home Human What's New BLAST Export Data Download Disease Browser Docs

Find All Lookup [e.g. AP000869, RH9632, cancer] Help

Flat File **FASTA** Gene List Feature List SNP List Image

Select data to export

Feature

☒ Gene ID: ENSG00000101200

Show context of: basepairs either side of feature

Region

Chromosome: Bands from: to

Contigs from: to

Markers from: to

Bases from: to

Select output format

☐ Text ☒ HTML ☐ Zip

Export Reset

Date: 2002-03-20 09:42:14 Help Desk / Suggestions

STEP 15

Choose the FASTA tab

STEP 16

Include 500 bp on either side, select HTML format

Ensembl Human Genome Server (ExportView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help Links

Ensembl Human ExportView The Wellcome Trust Sanger Institute EBI

Home Human What's New BLAST Export Data Download Disease Browser Doc

ENSG00000101200

>ENSG00000101200
 CAGACACTTGCATCGGGAATCCCGGGACACGGGTGGGTCCCTGAGCGGTTTGTGTTCAGT
 CTCCTGGGATGAAATGGGGACAGAATGGTCAACCCCTCGGCTGTCCCTCTGCCTCTCT
 CACTCTCTCTCCGCTACTGTCAAATGCTCTGCATCTCTCCAACTATCTGCTGTCCT
 GTCTCTGGGTCTCTTGTCTCTTGTCTCCAGGCTTTCCCTCTCTGCTTACTCTCTT
 CCGCTCTCGGGAGGATCTCAACCCACCCCTCCAAAGGCACCTCTGGGCTCTGGTCTC
 CTTCCGCTGCTCCCAATTGGCCCTGTCCCTCGGTCTAGCGACACCTCTGGGGTCCGA
 GATCGCTCTCTACGGGACTGTTGCTCTCTGGGGTGGATCTGCCACGACCTCAGGT
 GCGGGGACTGGCCGGTGGCGGTGTGGCGGGGGTCCGCGCTCCCACTCTCTCCCTTTC
 CTCTCTCCGCCAGAGACTGAGACAGACGCGAGGCGGTGCAATTGGCGGAGGTTATTGTC
 GTGCTGACGGGGCGGGCGGAAGAGCGCGCGGTGGGCGGAGCGCGGGCTCAGTAGGCG
 TGGGCTGGGCGGGCTCGAAGGGCTCGGGCGCCCGGCGCAGCTCGACCGCGCAGCAG
 AAGGCCCGCGCGCGCCCTCCAGCTGCTGCGCTTCTCCGCTGCTGCGCGGGCGGG
 CGGTGAAGGCCCTCGGGGCACTCGGCTCGGTCAAGCAGCTCTTGGCGGGGAGCGGTG
 GAGCAGGGGGGCTCTGGGGGGGCGAGCTCGGGGTGGCGGGGGGCCACACCTCTGCTG
 CGGGCGCGCAGCGCCCCACCCCGCGCAGGCGCGGCTCCGCCACACCAAGCGGTCTG
 CGGCGGCGCGCAGCGCGCGCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGG
 CGGGGCGAGCGGGGCTCCGCGAGCTTCTGGCGGCGAGTGCAGGGGCGAGCGGAGG
 TACTTCTCTCTCTGCGAGCGCAGCGCTCAGCGGTGCCAGGAGCGACCGAGCTCTC
 GACGAGAGATGCTGGGCGGAGGACGCGGCTTTGCCCCGGGGCGCAGGGAGGAC

STEP 17

Copy part of the sequence, then click on BLAST

Ensembl Human Genome Server (BLASTView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Links »

Ensembl Human BLASTView

The Wellcome Trust Sanger Institute EBI

Home Human What's New BLAST Export Data Download Display Browser Docs

Find All LookUp [e.g. AP000869, RH9632, cancer] Help

Ensembl BLAST Server

Alternative sequence search: [SSAHA](#)

RETRIEVE BLAST RESULTS Help

Enter the blast retrieval ID: Retrieve

SUBMIT A BLAST QUERY Help

Paste your sequence here in FASTA or plain text format.

Seq. Reset

OR select the sequence file you wish to search Browse

BLAST OPTIONS Help

Database: Homo sapiens, peptides (known protein)

Executable: BLASTX (transl. DNA vs. protein)

☒ Mask repetitive sequences using Repeatmasker.

☒ Filter low complexity regions.

☐ Display histogram of score statistics.

Report 100 alignments.

ADVANCED BLAST OPTIONS Help

Matrix: blosum62

Descriptions: 100

Sort results by: pvalue

Genetic Code: Standard (blastx only)

other options: (not validated)

Expect (E): 10

HSP score: sump

Filter type: seg

Date : 2002-05-07 13:42:58

[Help Desk / Suggestions](#)

STEP 18

Paste the sequence, choose the ‘peptides (known+novel)’ database and BLASTX, then click on Search

Ensembl Human Genome Server (BlastView) - Microsoft Internet Explorer

Ensembl Human BLASTView

Find: [All] [e.g. AP000869, RH9632, cancer] [Help]

Blast hits

AL160414.18.1.35836
p(n) 6.2e-36
Score: 300 ID: 100%

Blast score ranges for this search: [The highest] 43 - 128 128 - 214 216 Show in ContigView

Location of Blast hits

View standard format BLAST results

BLASTX 2.0a13MP-WashU [10-Jun-1997] [Build 23:08:19 Jun 10 1997]
Query: / (867 letters)
Database: ensembl/Homo_sapiens.pep.fa 29076 sequences; 12218578 total letters

	Sequences producing significant alignments:	High Score	Smallest Sum P(N)	Probability
20	ENSP00000217172 Gene: ENSG00000101200 Clone: AL12...	300	6.2e-36	2
15	ENSP00000258888 Gene: ENSG00000136383 Clone: AC01...	82	0.00046	2
17	ENSP00000007699 Gene: ENSG00000006047 Clone: AC00...	65	0.071	2
22	ENSP00000291060 Gene: ENSG00000100139 Clone: AL02...	67	0.11	3
19	ENSP00000221214 Gene: ENSG00000075702 Clone: AC00...	72	0.13	2
20	ENSP00000217386 Gene: ENSG00000101405 Clone: AL16...	75	0.14	1
19	ENSP00000284752 Gene: ENSG00000154545 Clone: AC02...	72	0.19	1

View the alignment

STEP 19
Display the region in ContigView

Ensembl Human Genome Server (ContigView) - Microsoft Internet Explorer

Detailed View

Jump to chr: [20] bp [2993630] to [3043630] [Refresh]

Zoom: [2 Mb] [1 Mb] [Window] [1 Mb] [2 Mb]

Features: DAS Sources: Decorations: Export: Jump to: Image size: Help

Length: 5,994 Mb to 5,998 Mb

RefSeq: [Gene List] [Flat file] [FASTA] [SNP List] [Image]

NCBI Transcr.: [Close menu]

Mouse matches: [Close menu]

Human mRNAs: [Close menu]

Unigene: [Close menu]

Proteins: [Close menu]

GeneScape: [Close menu]

EST Transcr.: [Close menu]

Transcripts: [Close menu]

Sanger Transcr.: [Close menu]

BLAST hits: [Close menu]

DNA(contigs): [Close menu]

BLAST hits: [Close menu]

Sanger Transcr.: [Close menu]

Transcripts: [Close menu]

EST Transcr.: [Close menu]

GeneScape: [Close menu]

Proteins: [Close menu]

Unigene: [Close menu]

Human mRNAs: [Close menu]

NCBI Transcr.: [Close menu]

RefSeq: [Close menu]

Gene Legend: SANGER CURATED KNOWN GENES SANGER CURATED NOVEL CDS SANGER CURATED PUTATIVE
SANGER CURATED NOVEL TRANS SANGER CURATED PROTEIN GENES
ENSEMBL PREDICTED GENES (KNOWN) ENSEMBL PREDICTED GENES (NOVEL)

Date: 2002-03-20 10:09:21 Help Desk / Suggestions

STEP 20
Zoom out, then export a Gene List

View the BLAST hits

Ensembl Human Genome Server (ExportView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Ensembl Human ExportView The Wellcome Trust Sanger Institute EBI

Home Human What's New BLAST Export Data Download Disease Browser Docs

Find All [e.g. AP000869, RH9632, cancer]

Flat File FASTA Gene List Feature List SNP List Image

Select genes to export

Region

Chromosome:

☐ Bands from: to

☐ Contigs from: to

☐ Markers from: to

☒ Bases from: to

☐ Entire Chromosome

☐ Entire Genome

Restrict selection

Include: ☐ Known genes only ☐ Exclude Known genes ☒ Both

Include: ☐ Disease genes only ☐ Exclude Disease genes ☒ Both

Include: ☐ Transmembrane domains only ☐ Exclude Transmembrane domains ☒ Both

Include: ☐ Low-complexity domains only ☐ Exclude Low-complexity domains ☒ Both

Include: ☐ Signal domains only ☐ Exclude Signal domains ☒ Both

Include only those genes with Family ID:

Include Gene data

Include the following data

☒ Chromosome

☒ Chromosome start

☒ Chromosome end

☒ Chromosome band

☐ Chromosome strand

☒ Ensembl gene ID

☐ Known gene external id

☐ Known gene external database n

☐ Contig start

☐ Contig end

☐ Contig strand

☐ Contig name

☐ FPC contig name

☐ Gene description

☐ Ensembl peptide ID

☐ InterPro ID

Chromosomal coordinates are filled in automatically

STEP 21

Choose an output format

Ensembl Human Genome Server (ExportView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

☐ InterPro ID

☐ InterPro description

☐ InterPro short description

☐ PFAM ID

☐ PRINTS ID

☐ Prosite ID

☐ Family ID

☐ Family description

Include Disease data

☐ Disease description

☐ Disease OMIM id

Include SNP data

SNP restrictions:

☒ All SNPs From (5'BP) To (3'BP)

☐ Coding only

☐ Intronic only

☐ 5' UTR only

☐ 3' UTR only

☐ 5' Upstream only From (5'BP)

☐ 3' Downstream only To (3'BP)

Include SNP info:

☐ Reference SNP id

☐ SNP class

☐ Validated

☐ SNP Chromosome strand

☐ Start (chromosome coordinates)

☐ Start (contig coordinates)

☐ Start (clone coordinates)

☐ Allele

☐ Mapweight

☐ Sub SNP count

☐ Contig

☐ TSC ID

☐ HGBASE ID

☐ start (transcript coordinates)

☐ start (coding sequence coordinates)

☐ SNP type

☐ Ensembl transcript ID

Select output format

☐ Text ☐ HTML ☐ Zip ☒ Excel

Output in Excel format

http://www.ensembl.org/Homo_sapiens/exportview - Microsoft Internet Explorer

File Edit View Insert Format Tools Data Go To Favorites Help								Links »		
A1								■ Gene data		
A		B		C		D		E	F	G
1 Gene data		Location on Chromosome								
2	Ensembl gene ID	External gene ID	Chromosome	Chromosome start	Chromosome end	Chromosome band	Chromosome strand			
3	ENSG00000149488	Q9BYN4	20	2530478	2546251	p13	1			
4	ENSG00000149439	Q9BYN5	20	2548847	2571657	p13	1			
5	ENSG00000101361	NOL5A	20	2588491	2594046	p13	1			
6	ENSG00000101365	IDH3B	20	2594050	2599845	p13	-1			
7	ENSG00000088881	Q9P2A6	20	2628531	2695754	p13	1			
8	ENSG00000088882	CPXM	20	2729728	2736235	p13	-1			
9	ENSG00000101386	Q9NUB4	20	2750664	2751483	p13	1			
10	ENSG00000166958		20	2752211	2752663	p13	1			
11	ENSG00000132635	C20orf81	20	2770978	2776339	p13	-1			
12	ENSG00000132670	VPS16	20	2776380	2802385	p13	1			
13	ENSG00000037980	NM_002836	20	2809356	2974007	p13	1			
14	ENSG00000125787	GNRH2	20	2979275	2981398	p13	1			
15	ENSG00000125901	Q9H273	20	2981711	2983825	p13	1			
16	ENSG00000101405	OXT	20	3007273	3008169	p13	1			
17	ENSG00000101200	AVP	20	3018193	3021065	p13	-1			
18	ENSG00000125819	Q9H4J2	20	3045263	3095542	p13	-1			
19	ENSG00000088899	NM_014731	20	3098280	3104214	p13	-1			
20	ENSG00000101205	C20orf116	20	3126019	3140302	p13	-1			
21	ENSG00000125877	ITPA	20	3145101	3159513	p13	1			
22	ENSG00000088836	SLC4A11	20	3163070	3173717	p13	-1			
23	ENSG00000088854	Q9UFX9	20	3186799	3343211	p13	-1			
24	ENSG00000088812	ATRN	20	3406762	3582498	p13	1			
25										
26										
27										
28										
29										

Ensembl Data Disclaimer /

Unknown Zone

Output in HTML format

Ensembl Human Genome Server (ExportView) - Microsoft Internet Explorer

Ensembl Human ExportView The Wellcome Trust Sanger Institute EBI

Home Human What's New BLAST Export Data Download Disease Browser Docs

Chromosome 20 2518630 to 3518630

gene_name	external_id	chromosome	chrom_start	chrom_end	chrom_band
ENSG00000149488	Q9BYN4	20	2530478	2546251	p13
ENSG00000149439	Q9BYN5	20	2548847	2571657	p13
ENSG00000101361	NOL5A	20	2588491	2594046	p13
ENSG00000101365	IDH3B	20	2594050	2599845	p13
ENSG00000088881	Q9P2A6	20	2628531	2695754	p13
ENSG00000088882	CPXM	20	2729728	2736235	p13
ENSG00000101386	Q9NUB4	20	2750664	2751483	p13
ENSG00000166958		20	2752211	2752663	p13
ENSG00000132635	C20orf81	20	2770978	2776339	p13
ENSG00000132670	VPS16	20	2776380	2802385	p13
ENSG00000037980	NM_002836	20	2809356	2974007	p13
ENSG00000125787	GNRH2	20	2979275	2981398	p13
ENSG00000125901	Q9H273	20	2981711	2983825	p13
ENSG00000101405	OXT	20	3007273	3008169	p13
ENSG00000101200	AVP	20	3018193	3021065	p13
ENSG00000125819	Q9H4J2	20	3045263	3095542	p13
ENSG00000088899	NM_014731	20	3098280	3104214	p13
ENSG00000101205	C20orf116	20	3126019	3140302	p13
ENSG00000125877	ITPA	20	3145101	3159513	p13
ENSG00000088836	SLC4A11	20	3163070	3173717	p13
ENSG00000088854	Q9UFX9	20	3186799	3343211	p13
ENSG00000088812	ATRN	20	3406762	3582498	p13

Date : 2002-03-20 10:34:15 Help Desk / Suggestions

Internet

Tasks

1. Exploring features related to a gene

Find the gene report for the human MDM2 gene (MDM2_human).

How many transcripts are there predicted for this gene? What is the size of the longest predicted mRNA? How many exons does it comprise? How many amino acids does it code for?

With which protein does the MDM2 protein interact? In which species was MDM2 first identified?

Which Interpro domains does MDM2 contain?

On which chromosomal band is MDM2 located? On which contig in the golden path?

Does MDM2 match any mouse traces?

2. Exploring a region

Display the region between markers D6S1993 and D6S2096 in ContigView (*Hint: start from chromosome 6 on the homepage*).

Centre the display around the PRDM1 gene by clicking on this gene in “Overview”. In “Detailed View”, why are some genes above and some below the DNA contigs?

What is the nearest marker to the PRDM1 gene? How many synonyms does this marker have?

Zoom out to the 1Mb view and export a gene list for this region. (*Hint: Use the drop-down menu at the top of “Detailed View”*). Include Ensembl gene id, known gene external id, known gene external database name, gene description, InterPro short description.

Export a SNP list for chromosomal band 6q21.

3. Examining the supporting evidence for a gene prediction

Display the ENPP3 gene (or your own favourite gene) in ContigView.

Examine the evidence for the gene prediction. (*Hint: Collapse the chromosome view and overview; compare the collapsed and expanded view of the supporting evidence, e.g. Unigene – Note that max. 7 tracks can be displayed for each section; if there are more entries in the database the display may change on multiple re-loads.*)

Tick 'Acembly Transcripts', 'Ensembl RefSeq', 'NCBI Genomescan' and 'NCBI Transcripts' under 'DAS sources' on the yellow bar at the top of 'Detailed View'. Compare the Ensembl gene prediction to these gene prediction models.

Click on the ENPP3 transcript (or your own gene) and examine the links / alignments given for this gene in GeneView. Note especially the RefSeq and SpTrEMBL entries. What do the Target %id and Query %id figures indicate? (*Hint: Consult the help page by clicking on 'Help' in the top right corner.*)

Follow the link 'View supporting evidence' for the transcript ENST00000263052. From there follow the link 'View Evidence' to TransView. Examine the evidence from the different databases.

4. Exploring the mouse genome with Ensembl

Hint: Use what you know about navigation in human Ensembl.

Find the 'generic' Ensembl homepage and select mouse. Bring up a ContigView display of chromosome 11 between 111.5 Mb and 111.6 Mb.

How many 'known' and 'novel' genes are predicted in this region? For one of the known genes, find some information about its function, and look at an entry for it in the MGI resource.

Can you find out anything about the possible function of the novel gene?

What supporting evidence was used to predict these genes?

Export a text file showing the gene id numbers, descriptions and chromosomal positions of the genes in this region.

Turn on the 'Tile Path' track to see BAC clones that have been mapped to the assembly. What is the difference between the blue bar in the middle of detailed view (labelled DNA (contigs)) - and the blue bar at the bottom (labelled Assembly Ctgs) (*Hint: bring up the help page for ContigView, and find the link to a page of mouse-specific information*).

Which human chromosome has a region of conserved synteny with this region of mouse chromosome 11? (*Hint: Jump to Syntenyview using the menus on the gold bar*). More about inter-species comparison is coming up later in the workshop.

Task answers

1. Enter the gene name into the text search box and follow the link to GeneView. The MDM2 gene has 8 transcripts. Click on the longest transcript. The length of the transcript is 1473 bp, 10 exons. Follow the 'View protein' link. The transcript codes for 491 aa. Follow the link from GeneView to Locuslink or OMIM. MDM2 was first identified in mouse, the protein binds to p53. Check the InterPro section in GeneView. The domains are ATP/GTP-binding site motif A (P-loop), RING finger, Zn-finger in ranbp and others, p53-associated protein (MDM2). Follow the link 'Genomic location'. MDM2 is located on 12q15. It is located on contig AC025423.32.1.150579 (see GeneView, Genomic Location, or mouse-over the contig in ContigView). Make sure mouse matches are ticked under 'Features' in ContigView. Yes, MDM2 matches mouse traces.

2. Start on the homepage and click on chromosome 6. Enter the markers in MapView. The genes above the contig track are in the forward orientation, the genes below are in the reverse orientation. Get marker information by mouse-over the closest marker. It is RH16128 (D6S1217E) and has 5 synonyms (Hint: zoom in first in order to distinguish single markers). Choose gene list from the drop-down menu at the top of 'Detailed View'. Enter the required data, chromosomal coordinates are filled in automatically. Choose the tab 'SNP list' and enter the chromosomal band plus the details of your choice.

3. No standard answer.

4. The quickest way to reach the region is to enter the coordinates in the browse chromosome boxes (you can use M or Mb) on the mouse homepage (http://www.ensembl.org/Mus_musculus/). There are 3 known genes and 1 novel gene. Get functional information via the brief description on the GeneView page and via links (e.g. to SwissProt and MGI). The protein and mRNA tracks in ContigView show the supporting evidence (each has a link to a database entry). Export using the pull-down menu on the gold bar. Turn on tile path from the Decorations pull-down menu on the gold bar. Click the red help icon in the top right to see the ContigView help page, and then look at the 'extra information on mouse displays'. The mouse sequence is an assembly of whole genome shotgun reads, not sequence from BAC clones. The WGS assembly contigs are shown at the bottom. For easier processing and annotation, Ensembl is presenting the final genome assembly in 5 Mb chunks.

Glossary

Golden Path term sometimes used for a genome-wide assembly of sequence contigs to make a single base sequence, including gaps, for each chromosome

Contig A stretch of contiguous (continuous) something. Confusingly, used in many different contexts. Most often, it means a stretch of continuous DNA sequence. But can also be used to mean a continuous assembly of fragments, or of clones, without implying that their sequence is known

NCBI National Centre for Biotechnology Information. Part of the U.S. National Library of Medicine (NLM), National Institutes of Health (NIH).

EBI European Bioinformatics Institute, Hinxton. An outstation of the European Molecular biology Laboratory.

URLs

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

Map Viewer at NCBI http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/hum_srch

Human Genome Browser at UCSC <http://genome.ucsc.edu>