

e!

Ensembl

Distributed Annotation System

DAS

DAS Specification

- Distributed Annotation System (DAS)
 - A server system for the sharing of Reference Sequences, and their annotations.
- Reference Sequence
 - A sequence.
- Annotation - An entity which:
 - Is anchored to the reference sequence via a stop and start value.
 - Possesses an ID unique to the server and a structured description of its types, methods, and categories.
- Global Annotation:
 - Annotation that applies to the entire reference sequence; start and stop are void

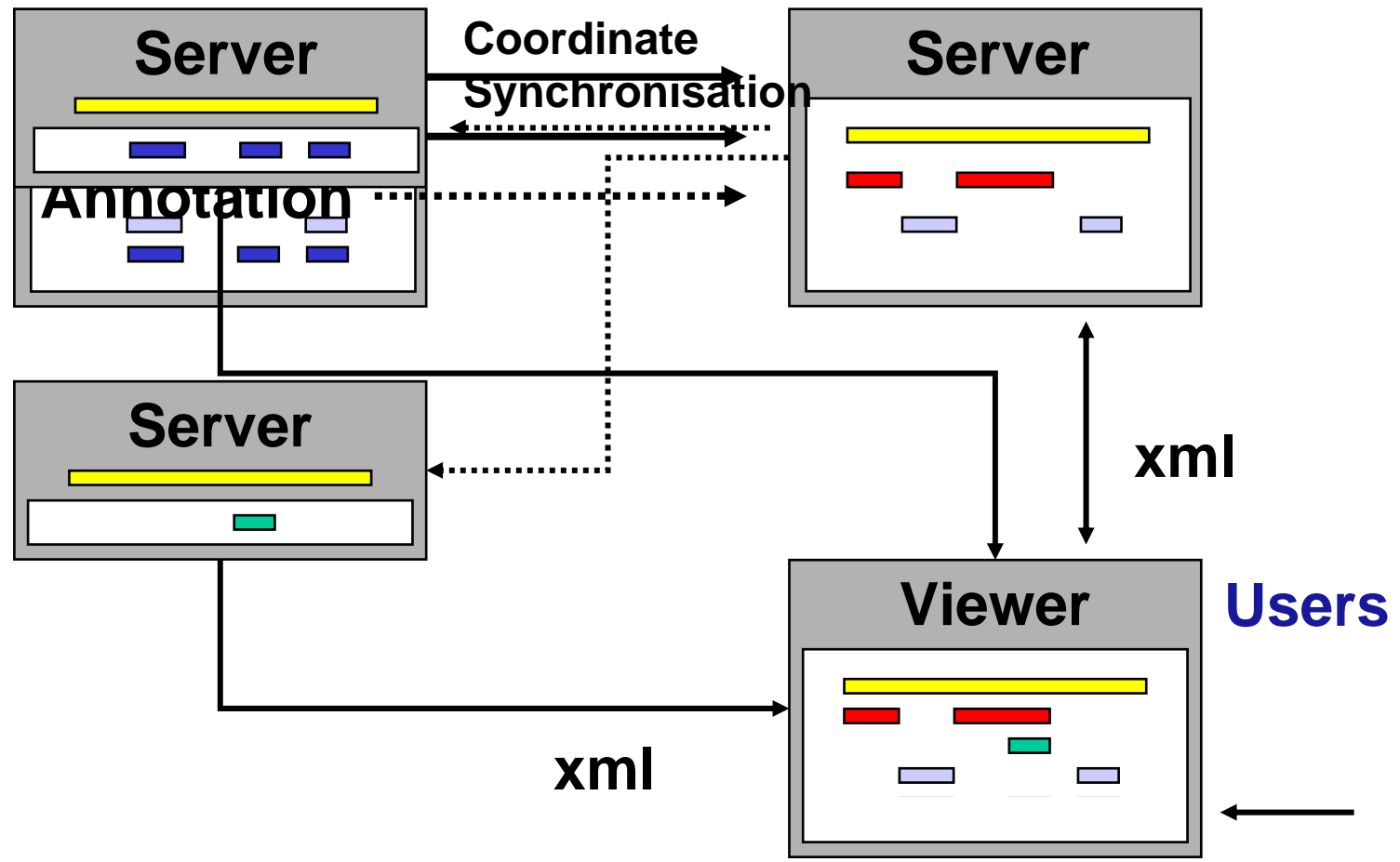
Reference Sequences

- Applicable to ContigView;
 - Chromosome
 - Contig
 - Clone
- Applicable to ProtView or GeneView;
 - Protein
 - Gene (global annotation only)
 - cDNA (not yet implemented)

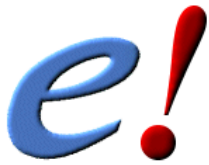
Distributed Annotation

External Contributors

Database providers

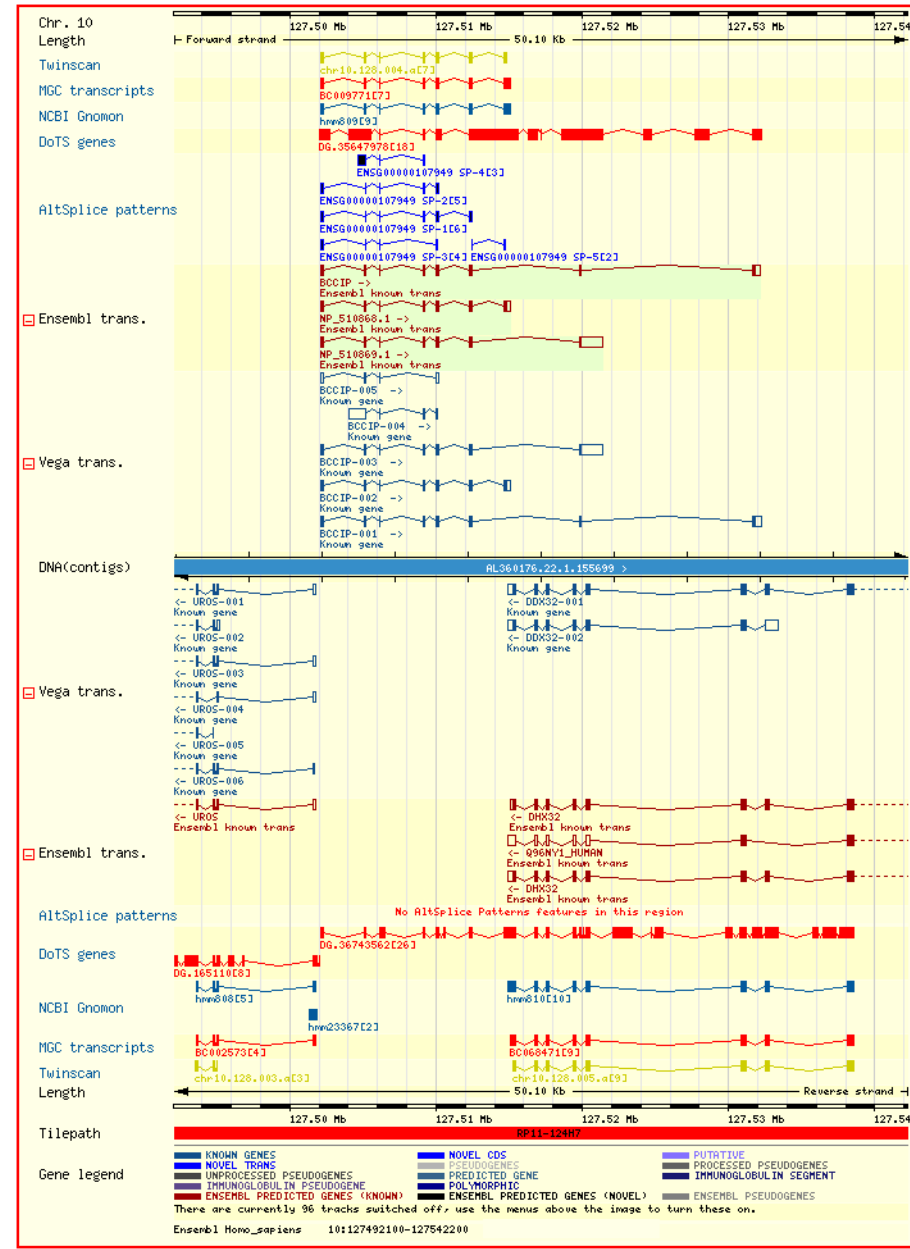


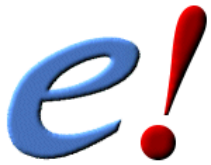
ensembl



ensembl

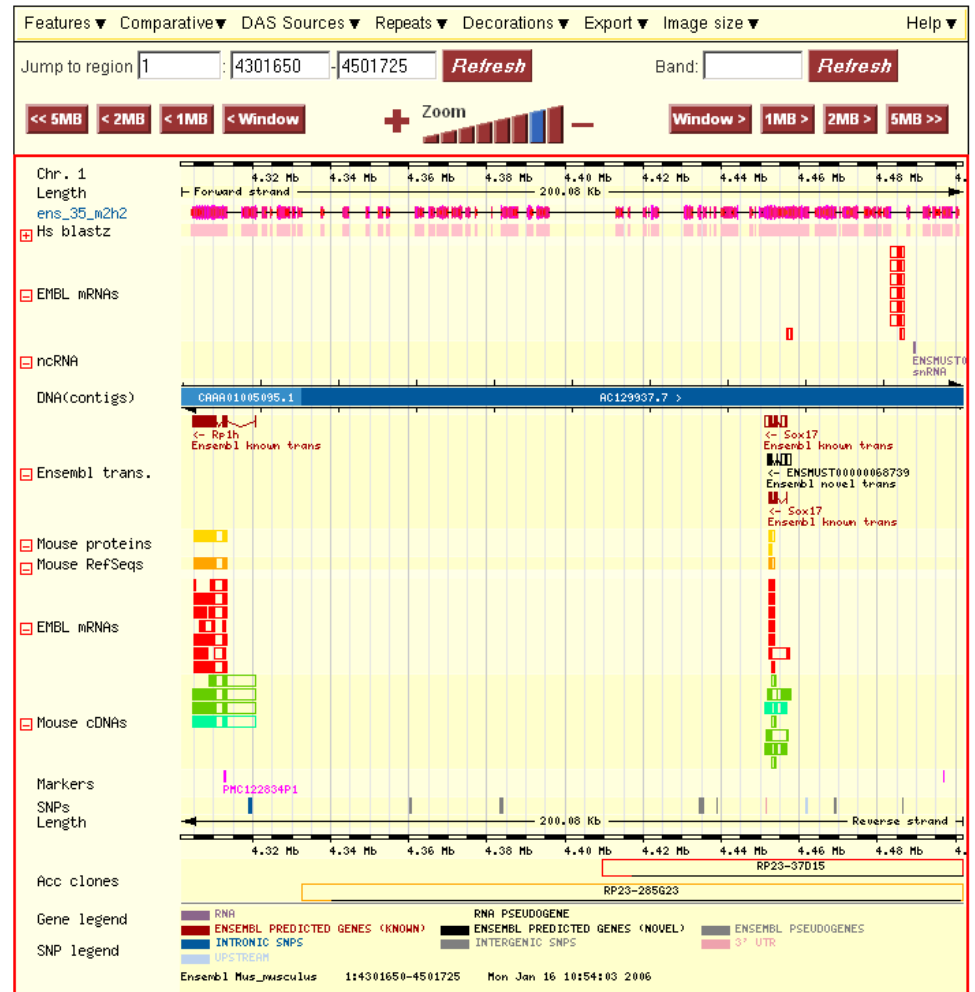
Display 3rd party data

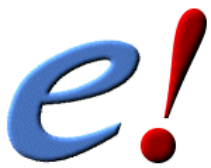




ensembl

Display Custom tracks





Display 3D structures

ensembl

Features ▾ Comparative ▾ DAS Sources ▾ Repeats ▾ Decorations ▾ Export ▾ Image size ▾ Help ▾

Jump to region 4 : 140182400 - 140210800 Refresh Band: Refresh

<< 5MB < 2MB < 1MB < Window + Zoom - Window > 1MB > 2MB > 5MB >>

Chr. 4
Length 140.19 Mb 140.19 Mb 140.19 Mb 140.20 Mb 140.21 Mb 140.21 Mb
28.40 Kb

Forward strand

PDB_Spice

EMBL mRNAs

Mouse proteins

Genscan

EST trans.

Ensembl trans.

Vega trans.

DNA(contigs)

PDB_Spice

Length

Gene legend

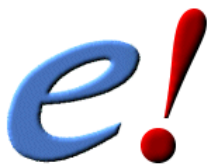
KNOWN PROTEIN CODING
EST GENES
ENSEMBL PREDICTED GENES (KNOWN) ENSEMBL PREDICTED GENES (NOVEL) ENSEMBL PREDICTED GENES (NOVEL)
There are currently 67 tracks switched off, use the menus above the image to turn them on.

Ensembl Mus_musculus 4:140182400-140210800 2006

PDB: 1mqb.A via UniProt P29317 PDB: EPHRIIN TYPE-A RECEPTOR 2

ID: ENSP - PDB mapping to 1mqb.A
TYPE: ENSP - PDB mapping
METHOD: Compara, MSD-Mapping of UniProt to PDB
SCORE: 96
FEATURE LOCATION:
- Start: 779
- End: 888
- Strand: Forward
DAS LINK:
<http://das.sanger.ac.uk/registry/showspice.jsp?pdb=1mqb.A>
NOTES: PDB: 1mqb.A 96%id 27%coverage via UniProt P29317 PDB:EPHRIIN TYPE-A RECEPTOR 2

run SPICE



Display 3D structures

Ensembl

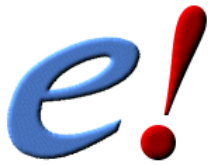
The screenshot displays the SPICE (Sequence-Profile-Integrated Comparison Environment) web interface. The main window is titled "SPICE" and has a menu bar with "File", "Display", and "Browse".

The central panel shows two 3D protein structures. The top structure is a ribbon representation of a protein dimer, colored in shades of green and yellow. The bottom structure is a similar ribbon representation, colored in shades of white and grey.

On the right side, there are several panels for sequence and domain annotations:

- PDB**: A horizontal bar representing the protein sequence from residue 1 to 263. Below it, a legend lists various structural motifs: dssp (BRIDGE, SECSTRUC, BEND, SHELV), cath (Cath Domai), and s3dm (ALPHA-BEL, ALPHA-BEL, BETA-BULG, BETA-BULG, BETA-TURN, BETA-TURN, BETA-TURN, GAMMA-TU, NEST-UP, NEST-RL, SCHELLMA, ST-MOTIF, ST-STAPLE).
- UniProt**: A horizontal bar representing the protein sequence from UniProt ID P29317. Below it, a legend lists various functional and structural annotations: description, SIGNAL, CHAIN, TOPO_DOM, CARBOHYD, COMPBIAS, DOMAIN, TRANSMEM, MOD_RES, SECSTRUC, NP_BIND, BINDING, ACT_SITE, MOTIF, Sim4, Scop, PROSITE, ProDom, PRINTS, and Pfam.
- ENSP**: A horizontal bar representing the protein sequence from Ensembl ID ENSP00000166244. Below it, a legend lists various annotations: ensp_pdb_mapping, MisPred, and hsa35pep.

At the bottom of the window, there is a text input field with the placeholder text "enter RASMOL like command...".



Ensembl

GeneDAS

ENSMUSG00000006445 **Ensembl Gene Report for ENSMUSG00000006445**

- Gene information
- Gene splice site usage
- Gene variation info.
- Genomic sequence
- Transcript information
- Exon information
- Protein information
- Export data

Chromosome 4
140,192,502 - 140,210,540

View of Chromosome 4

Gene	Epha2 (MGI Symbol) (Click here all Ensembl genes linked to the name (link text))
Ensembl Gene ID	ENSMUSG00000006445
Genomic Location	This gene can be found on Chromosome 4 at location 140,192,502-140,210,540 . The start of this gene is located in Contig AL607887.10
Description	Eph receptor A2 [Source:MakerSymbol;Acc:MG195270]
Prediction Method	Genes were annotated by the Ensembl automatic analysis pipeline using either a GeneWise/Exonerate model from a database protein or a set of aligned cDNAs followed by an ORF prediction. GeneWise/Exonerate models are further combined with available aligned cDNAs to annotate UTRs (For more information see V.Curwen et al., Genome Res. 2004 14 542-50.)
Transcript	ENSMUST00000008614 ENSMISF00000006614 Epha2 [Transcript info] [Exon info] [Exon info]

Gene DAS Report

DAS Sources

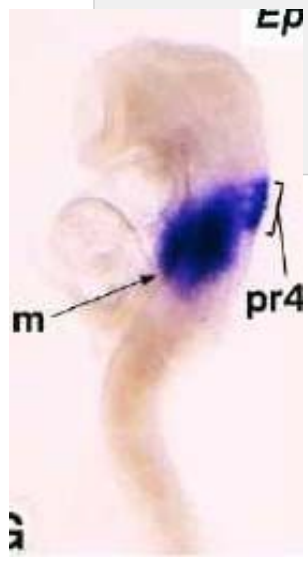
- [AltSplice](#) (Alternative Splice Database)
- [AltTrans](#) (Alternative Transcript Diversity Database)
- [EMMA](#) (European Mouse Mutant Archive)
- [Emage](#) (Mouse Embryo Spatial Gene Expression Database)
- [GeneTrap](#) (IGTC gene trap sequence mapping)
- [MGI_Phenotype](#) (MGI Phenotypic Alleles)
- Phenotypes (Associated directly or via orthologues or protein families)
- [Protonet](#) (Global classification of proteins into hierarchical clusters)
- [UniProt](#) (Protein knowledgebase)

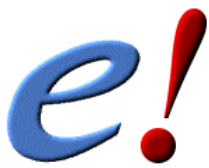
[Manage Sources](#)

Transcript	Epha2 (MGI Symbol) (Click here all Ensembl genes linked to the name (link text))
Transcript information	Exons: 17 Transcript length: 3,307 bps Protein length: 977 residues [Further Transcript info] [Exon information] [Protein information]
Similarity Matches	This Ensembl entry corresponds to the following database identifiers: MGI Symbol: Epha2 UniProt/Swiss-Prot: EPA2_MOUSE (Target No: 99, Query No: 99) [info] RefSeq protein: NP_034269.2 (Target No: 100, Query No: 100) [info] RefSeq EMBL: NM_010139.2 (Target No: 100, Query No: 100) [info]

GeneDAS

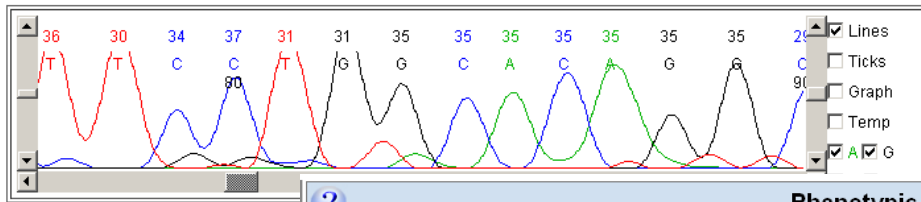
Emage	description Epha2 description expression data in Theiler Stage [12] in [9] Tissues
MGI_Phenotype	Phenotypic Allele MGI:2448472 MGI SYMBOL: Epha2[Gt(ROSABetageo)CN3Yiw] NAME: ge PUBMED: [11287184] Phenotypic Allele MGI:2668054 MGI SYMBOL: Epha2[Gt(U3Betageo)J3A3RuI] NAME: gene [8649815] Phenotypic Allele MGI:3040415 MGI SYMBOL: Epha2[tm1Jinc] NAME: targeted mutation 1, .
<input type="checkbox"/> DAS Sources	<input type="checkbox"/> AltSplice (Alternative Splice Database) <input type="checkbox"/> AltTrans (Alternative Transcript Diversity Database) <input type="checkbox"/> EMMA (European Mouse Mutant Archive) <input checked="" type="checkbox"/> Emage (Mouse Embryo Spatial Gene Expression Database) <input type="checkbox"/> GeneTrap (IGTC gene trap sequence mapping) <input checked="" type="checkbox"/> MGI_Phenotype (MGI Phenotypic Alleles) <input type="checkbox"/> Phenotypes (Associated directly or via orthologues or protein families) <input type="checkbox"/> Protonet (Global classification of proteins into hierarchical clusters) <input type="checkbox"/> UniProt (Protein knowledgebase) <input type="button" value="Manage Sources"/>





ensembl

http://tikus.gsf.de/project/sequence/19925q.html
G079C04 - 19925



1 - 10:	0	0	0	0	4	6
11 - 20:	19	14	14	9	9	10
21 - 30:	9	12	19	25	23	11
31 - 40:	6	6	6	8	17	18
41 - 50:	9	9	12	23	23	29
51 - 60:	25	29	22	22	22	22
61 - 70:	35	32	29	29	29	29
71 - 80:	33	28	26	26	26	26
81 - 90:	42	44	42	37	37	42
91 - 100:	40	45	35	36	30	34
101 - 110:	35	35	35	35	35	35
111 - 120:	19	28	19	24	19	23
121 - 130:	30	29	24	29	28	28
131 - 140:	26	28	26	26	33	33
141 - 150:	11	23	23	35	42	42
151 - 160:	35	29	33	33	39	37
161 - 170:	37	37	37	35	35	35
171 - 180:	33	37	37	44	35	35
181 - 190:	42	42	35	35	35	35
191 - 200:	35	35	32	30	30	33
201 - 210:	32	32	33	32	42	44
211 - 220:	42	42	40	30	30	27
221 - 230:	29	29	29	37	42	42
231 - 240:	56	56	56	56	56	56
241 - 250:	56	56	56	56	56	56
251 - 260:	48	48	37	21	17	4
261 - 270:	11	10	10	10	12	10
271 - 280:	27	33	30	16	13	10
281 - 290:	16	16	16	19	21	21
291 - 300:	0	0	0	0	0	0
301 - 305:	0	0	0	0	4	

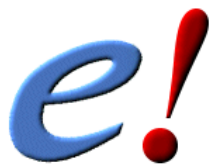
Phenotypic Allele Detail Your Input Welcome

Allele	Symbol: Epha2^{Gt(U3Betageo)J3A3RuI} Name: gene trap J3A3, H Earl Ruley ID: MGI:2668054					
Synonyms	eck1					
Allele details	Allele Type: Gene trapped Strain of Origin: 129/Sv ES Cell Line: Not Specified ES Cell Line Strain: Not Specified Mutation: Disruption caused by insertion of vector A provirus, containing Betageo, inserted into an intron 1.8 kb downstream of the alternatively spliced exon 5.2. Endogenous protein was undetected in homozygous mutant mice by Western blot analysis. (?31973) International Mouse Strain Resource: (Search for IMSR strains with Epha2 mutations) References and Additional Notes: (See Below)					
Gene information	Symbol: Epha2 Name: Eph receptor A2 Chromosome: 4 Genetic Position: 73.2 cM, cytoband D-E Genome Coordinates: 140182502-140210640 bp, + strand (From NCBI annotation of NCBI Build 34) Human Ortholog: EPHA2					
Phenotypes	Phenotypic details for all genotypes that include at least one Epha2^{Gt(U3Betageo)J3A3RuI} allele <table border="1"> <thead> <tr> <th>Allelic Composition</th> <th>Genetic Background</th> </tr> </thead> <tbody> <tr> <td>Epha2^{Gt(U3Betageo)J3A3RuI}/Epha2^{Gt(U3Betageo)J3A3RuI}</td> <td>involves: 129/Sv * C57BL/6</td> </tr> </tbody> </table> <p>normal phenotype no phenotype detected (?31973)</p>		Allelic Composition	Genetic Background	Epha2^{Gt(U3Betageo)J3A3RuI}/Epha2^{Gt(U3Betageo)J3A3RuI}	involves: 129/Sv * C57BL/6
Allelic Composition	Genetic Background					
Epha2^{Gt(U3Betageo)J3A3RuI}/Epha2^{Gt(U3Betageo)J3A3RuI}	involves: 129/Sv * C57BL/6					
References	(Original) 1:31973 Chen J <i>et al.</i> , "Germ-line inactivation of the murine Eck receptor kinase by gene trap retroviral insertion." <i>Oncogene</i> 1996 Mar 7;12(5):979-88 All references(1)					

```

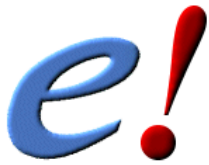
*****
GGGGGAAAAA
AAAAAAAAANN
NNNNNNNNNN
NNNNT

```



Adding custom data to Ensembl Displays

- Add a new DAS server;
 - Switch on pre-configured server,
 - Configure a new server,
 - Set up your own server (see Ensembl docs),
 - ContigView, ProtView and GeneView.
- Use Ensembl DAS server to display your data, made accessible via;
 - file upload,
 - Web page (URL),
 - ContigView only at this time.



Ensembl

Display of uploaded data

Features ▾ Comparative ▾ DAS Sources ▾ Repeats ▾ Decorations ▾ Export ▾ Image size ▾ Help ▾

Jump to region 32K BAC Rearray AceView Transcripts AltSplice SNPs AltSplice patterns BAC End Pairs

Band:

e! Ensembl Human HelpView

Ensembl v37 - Feb 2006 [Contact helpdesk](#)

- Help with help!**
- General
 - Full text search
- Entry Points**
- Ensembl BLASTView
 - Ensembl ExaSearch
 - Ensembl HelpView
 - Ensembl Homepage
 - Ensembl TextView
- Data Displays**
- Ensembl AlignSliceView
 - Ensembl AlignView
 - Ensembl ChromoView
 - Ensembl ContigView
 - Ensembl CytoView
 - Ensembl DiseaseView
 - Ensembl DomainView
 - Ensembl DotterView
 - Ensembl ExonView
 - Ensembl FamilyView
 - Ensembl FastaView
 - Ensembl FeatureView
 - Ensembl GeneSNPView
 - Ensembl GeneView
 - Ensembl GOView
 - Ensembl KaryoView
 - Ensembl LDView

URL based data.

Introduction

The 'URLSource' display allows for selection of custom data sets located on an external web server to be displayed in Ensembl 'ContigView' and 'CytoView' displays.

This system is based on the custom annotation track system of the [UCSC Genome Browser](#) and details about the data formats can be found at: <http://genome.ucsc.edu/goldenPath/help/customTrack.html>

To add your web-based track to Ensembl, you need to:

- Format the data set
- Define information about the track
- Choose your window of the genome
- Display in Ensembl
- Creating a URL to show your data

You may also find the following useful:

- Deleting tracks
- Publishing your data
- Example data

Formatting the Data Set

Firstly format your data as tab or space separated file using one of the following formats:

- GFF
- GTF
- BED
- PSL

Details of these formats can be found at <http://genome.ucsc.edu/goldenPath/help/customTrack.html>

Notes:

- You may include information for multiple tracks in the same file
- As entries in GFF/GTF files may contain spaces, then columns in GFF/GTF files MUST be separated by a tab or at least 2 spaces.

urlsource

8,000 9,000

GENSCAN000000007772
Ab-initio Genscan

AC097642.3.1.44638

Reverse strand

8,000 9,000

NC05-295222

ence for details.

Mozilla Firefox

data

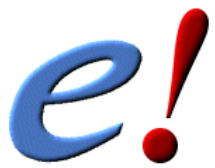
h a local web-based data-source to the Ensembl ContigView and CytoView displays

Date URL:

Fields marked with * are required

[SI / EBI](#). Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

URL based data...
CLOSE MENU ▲



URL-based

```
browser position chr2:1-10000
track name=Ensembl_test description="Ensembl
workshop (BED)" color=000000
url=http://www.ebi.ac.uk/~xose/ensembl_test.html
2      1000    1100    bed_feature_1 1000    +
2      2000    2100    bed_feature_2 500     +
2      3000    3100    bed_feature_2 100     +
```

Ensembl

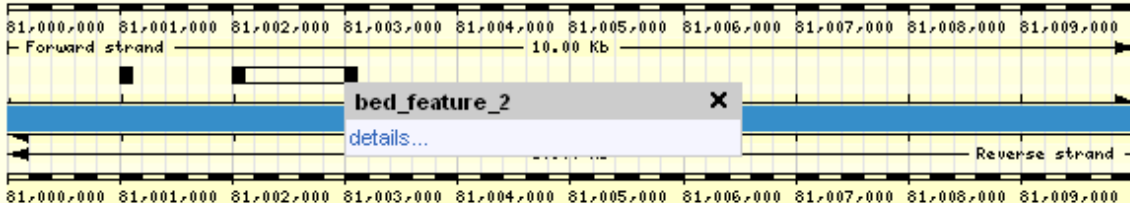
Display of data via URL

Features ▼ Comparative ▼ DAS Sources ▼ Repeats ▼ Decorations ▼ Export ▼ Image size ▼ Help ▼

Jump to region : - **Refresh** Band: **Refresh**

+
Zoom
-

Chr. 1
 Length
 Ensembl_test
 DNA(contigs)
 Length



81,000,000 81,001,000 81,002,000 81,003,000 81,004,000 81,005,000 81,006,000 81,007,000 81,008,000 81,009,000 8

Forward strand 10.00 Kb

bed_feature_2 X
[details...](#)

Reverse strand

81,000,000 81,001,000 81,002,000 81,003,000 81,004,000 81,005,000 81,006,000 81,007,000 81,008,000 81,009,000 8

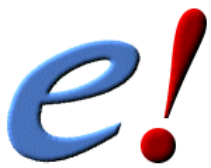
Tilepath

RPS-1189N14

RP1-54E18

There are currently 111 tracks switched off, use the menus above the image to turn these on.

Ensembl Homo_sapiens 1:81000000-81010000 Mon Mar 20 13:48:39 2006



Display of uploaded data

DAS Wizard Step 2 of 3: Data appearance

Coordinate System Provided by Registry

Enable on geneview protview transview
 contigview cytoview

[Back](#) [Next](#)

DAS sources

Name	DAS Server
das_32K	http://das1.s
das_ACEVIEW	http://das.er
das_ALTSPLICE	http://www.e
das_ALTSPLICESNP	http://www.e
das_Atlas	http://db.sys
das_AtlasPlasma	http://db.sys
das_BAC	http://das1.s
das_CHORI507	http://das1.s
das_CHORI_243	http://das.er
das_cisRED	http://island
das_COMPUGEN	http://das.er
das_CPG	http://das.er
das_Decipher	http://das.s
das_DOTS	http://das1.s
das_FIRSTEF	http://das1.s
das_FOSMID	http://das1.s
das_GCS_ExoFish	http://www.g
das_GCS_mRNA	http://www.g

DAS Wizard Step 3 of 3: Display configuration

Name: As nickname in registry

Label: As nickname in registry

Help URL:

Link Text:

Link URL:

Track colour:

Group features:

Display on:

Max rows to display:

Label features:

Apply stylesheet:

Use score:

[Back](#) [Finish](#)

DAS sources

Name	DAS Server	Data Source	Coordinate System
das_32K	http://das1.sanger.ac.uk:7070/perl/das	ens_ncbi_35_32karray	Ensembl Location
das_ACEVIEW	http://das.ensembl.org/das	ens_35_aceview	Ensembl Location
das_ALTSPLICE	http://www.ebi.ac.uk/das-srv/asd/das	asd_human	Ensembl Location
das_ALTSPLICESNP	http://www.ebi.ac.uk/das-srv/asd/das	asd_snp_human	Ensembl Location
das_Atlas	http://db.systemsbioology.net:8080/das	Human_Ens30_NCBI35_APD0405_P09	Ensembl Location
das_AtlasPlasma	http://db.systemsbioology.net:8080/das	HumanPlasma_ALL_Ens32_P09	Ensembl Location

ensembl

File Upload-based

```

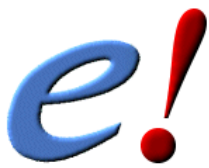
#<col1>          <col2>          <col3>          <col4>
#<group>         <name>          <type>         <subtype>

Similarity      Fake_match_1    homology       wublastn
Transcription   Fake_tscr_1     transcript     exon
Transcription   Fake_tscr_1     transcript     exon

#<col15> <col16>  <col17> <col18>  <col19> <col10>
#<chr>   <start> <end>   <strand><phase> <score>

2         4000    4050    +         .         100
2         4200    4300    +         .         100
2         4400    4500    +         .         100

```



ensembl

DAS registry



111110101010101
101101000101010
101010110100010
011001111000101

- [registry home](#)
- [list sources](#)
- [register new](#)
- [validate source](#)
- [statistics](#)
- [documentation](#)

available DAS services

Keyword search: search

organism: authority: type: capability: label:

Nucleotide sequences

European Bioinformatics Institute

[EBI Home](#) [About EBI](#) [Groups](#) [Services](#) [Toolbox](#) [Databases](#) [Downloads](#) [Submissions](#)

DATABASE BROWSING TOOLS

id	clients	nickname
DS_109		uniprot
DS_110		dssp
DS_111		cath
DS_112		structure
DS_113		alig_pdb_sp

- [UniProt DAS Home](#)
- [Access DAS Server](#)
- [Download DAS Server](#)
- [Contact Us](#)

UniProt DAS Services at the EBI

The distributed annotation system (DAS) is a client-server system in which a single client integrates information from multiple servers. The UniProt DAS server allows researchers to show their research results, for example identified peptides or signal sequences, on the UniProt reference sequence server, in the context of UniProt/Swiss-Prot annotation.



The UniProt Protein DAS Reference Server

- ◆ [Access the UniProt DAS Server](#)

The UniProt DAS Reference Server serves both sequence and feature data from UniProt/Swiss-Prot and UniProt/TrEMBL. In addition to this, feature links to InterPro for all the proteins in UniProt are provided, including basic information (position, id and a brief description) and internet addresses to the relevant data source.

The server can be queried using any of the following:

- ◆ *UniProt Accession numbers e.g. O35502 [sequence](#), [features](#)
- ◆ *Swiss-Prot ID's e.g. A4_Human [sequence](#), [features](#)
- ◆ *TrEMBL ID's e.g. Q12368 [sequence](#), [features](#)
- ◆ IPI ID's e.g. IPI00015171 [sequence](#)
- ◆ UniParc ID's e.g. UPI0000125656 [sequence](#)

All of these kinds of identifier will return sequence information. Identifier types marked with an * will also return UniProt and InterPro feature hits.

For further details of the UniProt DAS server click on the link above that will take you to the UniProt DAS home page. This includes example requests using the DAS protocol for real proteins that can be found in UniProt.

<http://das.sanger.ac.uk/registry/>

e!

Ensembl

QUESTIONS
ANSWERS