



Bioinformatics Tools and Resources for Identifying Gene Product Function



Emmanuel Dialynas

Bioinformatics Support Group @ IMBB



Proprietary -> Open Source

Command line -> Web

Web

Repositories/Databases

Data mining

Analysis

Sharing

Collaboration

Application Programming Interface

High Performance Computing

Paying tribute....



Genbank



DNA Database of
Japan



European
Nucleotide
Archive

Examples: [BN000065](#), [histone](#)

Search

[Advanced](#)
[Sequence](#)

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

Text Search

Examples: [BN000065](#), [histone](#)

Search

[Advanced search](#)

Sequence Search

Search

[Advanced search](#)

Popular

- [Submit and update](#)
- [Sequence submissions](#)
- [Genome assembly submissions](#)
- [Submitting environmental sequences](#)
- [Citing ENA data](#)
- [Rest URLs for data retrieval](#)
- [Rest URLs to search ENA](#)

Latest ENA news

02 Nov 2015: [Change to Globus endpoint for public ENA data](#)

The Globus endpoint for public ENA data is changing from `ebi#ena` to `ebi#public` ('ena' subfolder).

23 Sep 2015: [ENA Release 125](#)

Release 125 of ENA's assembled/annotated sequences now available

Examples: [BN000065](#), [histone](#)

[Advanced](#)
[Sequence](#)
[Home](#) [Search & Browse](#) [Submit & Update](#) [About ENA](#) [Support](#)

 Please subscribe to ena-announce mailing list here: listserv.ebi.ac.uk/mailman/listin... to receive alerts about ENA services.

Search results for *sodium channel*

Sequence

Sequence (Update) (5,877)

Sequence (Release) (55,612)

Coding

Coding (Update) (665)

Coding (Release) (6,349)

Study

Study (2)

Study (Sequence) (54)

Coding (Release) (6,349 results found)

 Download: - of 6,349 results in [TEXT](#) [XML](#) [FASTA](#)

Showing results 11 - 20 of 6,349 results

AAA16202	Doryteuthis opalescens sodium channel
AAA67105	Mus musculus (house mouse) sodium channel 21
AAA67106	Mus musculus (house mouse) sodium channel 25
AAP19643	Mus sp. partial sodium channel
ADO28451	Ictalurus punctatus (channel catfish) sodium channel modifier 1
AGO33659	Culex pipiens pallens sodium channel
BAA78033	Homo sapiens (human) sodium channel
BAM84088	Thrips palmi partial sodium channel
BAO52751	Thrips palmi sodium channel
BAO52752	Thrips palmi sodium channel

 Prev Next Go to page:

Coding: AGO33659.1

Culex pipiens pallens sodium channel

[Send Feedback](#) 

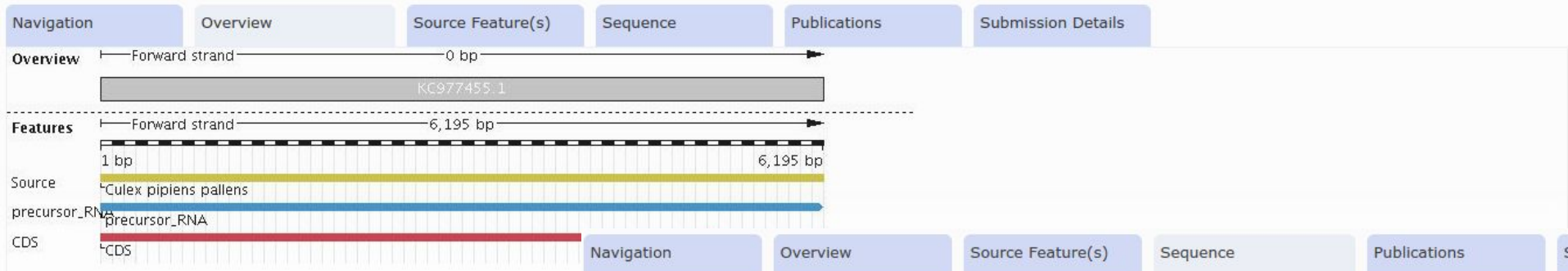
View: [TEXT](#) [FASTA](#) [XML](#)

Download: [XML](#) [FASTA](#) [TEXT](#)

Organism Culex pipiens pallens	Molecule type transcribed RNA	Topology linear	Data class STD	Taxonomic Division INV
Sequence length 6,132	Sequence Version 1			

Lineage

[Eukaryota](#), [Metazoa](#), [Ecdysozoa](#), [Arthropoda](#), [Hexapoda](#), [Insecta](#), [Pterygota](#), [Neoptera](#), [Endopterygota](#), [Diptera](#), [Nematocera](#), [Culicoidea](#), [Culicidae](#), [Culicinae](#), [Culicini](#), [Culex](#)



Showing first 1 - 1000 of 6132

```
>ENA|AG033659|AG033659.1 Culex pipiens pallens sodium channel : Location:1..1000
ATGACCGAAGACCTCGATTGATATCTGAGGAAGAACGTAGTTTGTCCGTCCTTTCACC
CGTGAATCATTACTTGTATTGAAGAACGCATCACAAATGAACAGGCAAAACAGCGCGAA
TTGGAAAAGAAGCGAGCCGAGGGGGAGATACGGTACGACGATGAGGACGAGGATGAAGGT
CCCCAGCCGGACTCCACACTCGAGCAGGGAGTGCCGATCCCTGTTTCAATGCAGGGCAGC
TTCCCTCCGGAATTGGCTCCACGCCTCTCGAGGATATTGACGCTTTTACTCAAACATT
AAGACATTCGTAGTAGTTAGTAAAGGAAAGGATATTTTCGTTTCTCGCAACCAATGCA
```

Want more?



Genome Browser

Gene Expression

Variance

Cross Species Comparison

Search: for

 e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [coronary heart disease](#)

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Popular genomes


Human

GRCh38.p3


Mouse

GRCm38.p4


Zebrafish

GRCz10

 ★ [Log in to customize this list](#)

All genomes

[View full list of all Ensembl species](#)

 Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#)

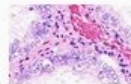
ENCODE data in Ensembl



Variant Effect Predictor



Gene expression in different tissues



Find SNPs and other variants for my gene

```

GTRFATACATTG
CRTRAAAGTCTT
CTTCTAAATTCT
GRAACATTTTCC
    
```

Retrieve gene sequence

```

GCGTGACTTCGGGGTGG:
GGGCTTGTGGGGGAGC'
GCGCCTCTGCTGGCCTT
AGGGGACAGATTTGTGA
CACCTCTGGAGCGGTTI
CCCAGTCCAGCGTGGCGI
    
```

Compare genes across species



Use my own data in Ensembl



Learn about a disease or phenotype



Did you know...?

"id": "!" Java, Perl, Python and Ruby, oh my! Try our [REST API](#) for "seq": ' quick access to Ensembl data.

'LPSSLSVI

Ensembl - Genome Browser

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Gene alleles
- Sequence
 - Secondary Structure
- External references
- Regulation
- 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

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gene: Scn2b ENSMUSG00000070304

Description sodium channel, voltage-gated, type II, beta [Source:MGJ Symbol;Acc:MGJ:106921 [↗](#)]

Location [Chromosome 9: 45,117,876-45,130,070](#) forward strand.
GRCm38:CM001002.2

About this gene This gene has 2 transcripts ([splice variants](#)), [56 orthologues](#), [4 paralogues](#) and is a member of [1 Ensembl protein family](#).

Transcripts [Show transcript table](#)

Summary [?](#)

Name [Scn2b](#) [↗](#) (MGI Symbol)

CCDS This gene is a member of the Mouse CCDS set: [CCDS40606.1](#) [↗](#)

UniProtKB This gene has proteins that correspond to the following Uniprot identifiers: [Q56A07](#) [↗](#)

RefSeq Overlapping RefSeq Gene ID [72821](#) [↗](#) matches and has similar biotype of protein_coding

Ensembl version ENSMUSG00000070304.10

Gene type Known protein coding

Annotation method Annotation produced by the Ensembl [genebuild](#).

[Go to Region in Detail for more tracks and navigation options \(e.g. zooming\)](#)

MGI transcript name: Scn2b-202

Gene	sodium channel, voltage-gated, type II, beta
	ENSMUSG00000070304
Location	Chromosome 9: 45,117,876-45,130,070
	↗ 🔍
Transcript	ENSMUST00000170998
	Exons
	cDNA Sequence
Protein	ENSMUSP00000126826
	Protein Variations
Gene type	Known protein coding
Transcript type	Known protein coding
Strand	Forward
Base pairs	4,562
Amino acids	215
Source	Ensembl

Genes (Comprehensive.....)

Contigs

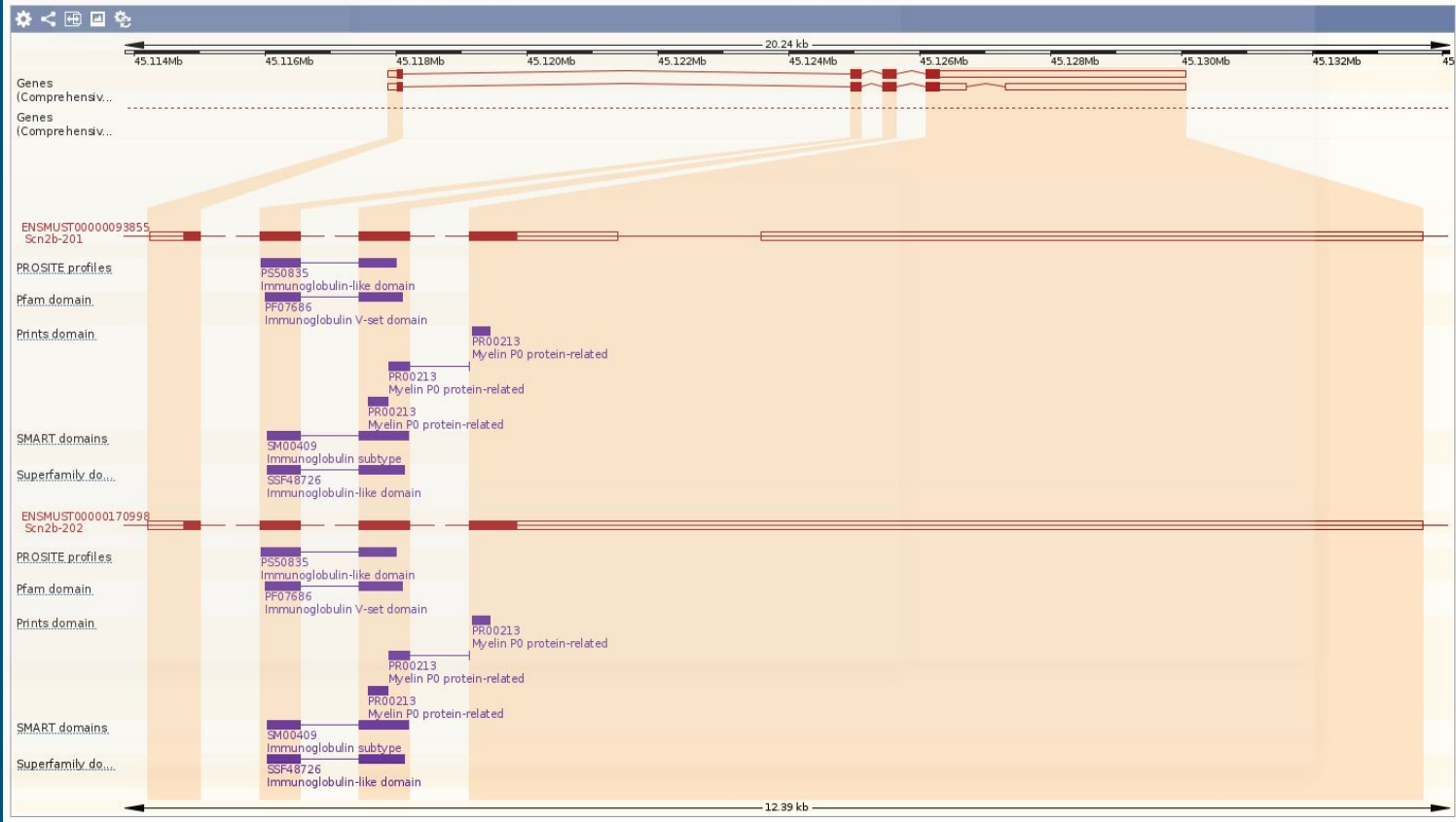
Genes (Comprehensive.....)

Gene_Legend

- Protein Coding
- Ensembl protein coding
- merged Ensembl/Havana

Ensembl - Genome Browser

Splice variants



Even more?



VectorBase
Bioinformatics Resource for
Invertebrate Vectors of Human Pathogens



Saccharomyces
GENOME DATABASE

WormBase



Ensembl - Genome Browser

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Sequence
- External references
- Regulation
- Expression report
- Pathways
- PubMed (8)
- Ontology
 - GO: molecular function
 - GO: cellular component
- Comparative Genomics
 - Genomic alignments**
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Phenotype
- Genetic Variation
 - Variation table
 - Structural variation
 - Variation image
- External data
 - Gene expression
 - Personal annotation
- ID History
 - Gene history

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gene: para AGAP004707

Description voltage-gated sodium channel [Source:VB Community Annotation;Acc:AGAP004707]

Synonyms VSC, kdr


Location [Chromosome 2L: 2,358,158-2,431,617](#) forward strand.

About this gene This gene has 3 transcripts ([splice variants](#)), [40 orthologues](#) and [9 paralogues](#).

Transcripts [Show transcript table](#)

Genomic alignments

Alignment:

 Download alignment

[View an image of this alignment](#)

A total of 7 alignment blocks have been found. Please select an alignment to view by selecting a Block from the Alignment column.

Show/hide columns		Filter	
Alignment (click to view)	Length (bp) ▾	Location on <i>Anopheles gambiae</i> ⚡	Location on <i>Anopheles coluzzii</i> ⚡
Block 1	17419	2L:2414259-2431677	scf_1925491148:12879-29794
Block 2	13277	2L:2394964-2408240	scf_1925489251:1-12378
Block 3	12410	2L:2375864-2388273	scf_1925488891:2-12424
Block 4	11838	2L:2358098-2369935	scf_1925490034:21918-31885
Block 5	3392	2L:2407946-2411337	scf_1925490032:1-3426
Block 6	2722	2L:2390790-2393511	scf_1925488609:1-2722
Block 7	1483	2L:2374139-2375621	scf_1925490009:1-1485

Ensembl - Genome Browser

GO: cellular component

Annotated terms

Ancestry chart

The following terms describe the *cellular_component* of this entry in [Gene Ontology](#)

Chart	Accession	Term	Evidence	Annotation Source	GOSlim Accessions	GOSlim Terms
	GO:0001518	voltage-gated sodium channel complex	IEA	UniProtKB/TrEMBL:Q7PJH0_ANOGA	GO:0005575 GO:0005623 GO:0043234 GO:0005886	cellular_component cell protein complex plasma membrane

GO: molecular function

Annotated terms

Ancestry chart

The following terms describe the *molecular_function* of this entry in [Gene Ontology](#)

Chart	Accession	Term	Evidence	Annotation Source	GOSlim Accessions	GOSlim Terms
	GO:0005248	voltage-gated sodium channel activity	IEA	UniProtKB/TrEMBL:Q7PJH0_ANOGA	GO:0003674 GO:0022857	molecular_function transmembrane transporter activity
	GO:0005509	calcium ion binding	IEA	UniProtKB/TrEMBL:Q7PJH0_ANOGA	GO:0003674 GO:0043167	molecular_function ion binding

Genome Browser Variation



Ensembl - Genome Browser

Transcript: para AGAP004707-RC

Description voltage-gated sodium channel [Source:VB Community Annotation;Acc:AGAP004707]

Synonyms VSC, kdr

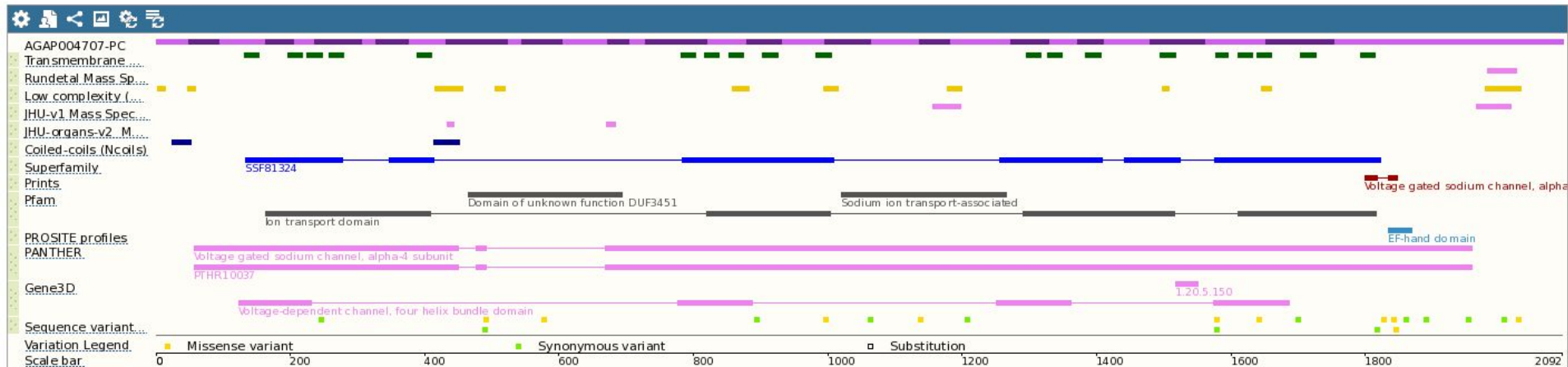
Location [Chromosome 2L: 2,358,158-2,431,617](#) forward strand.

About this transcript This transcript has [31 exons](#), is annotated with [72 domains and features](#), is associated with [22 variations](#) and maps to [40 oligo probes](#).

Gene This transcript is a product of gene [AGAP004707](#) [Show transcript table](#)

Protein summary ?

Protein domains for AGAP004707-PC.1



Statistics

- Ave. residue weight: 112.170 g/mol
- Charge: -55.5
- Isoelectric point: 4.7694
- Molecular weight: 234,659.09 g/mol
- Number of residues: 2,092 aa

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB

UniProt
Knowledgebase

Swiss-Prot
(549,832)

Manually
annotated and
reviewed.

TrEMBL
(54,540,801)

Automatically
annotated and not
reviewed.

UniRef

Sequence clusters



UniParc

Sequence archive



Proteomes



Supporting data

Literature citations



Cross-ref. databases



Taxonomy



Diseases

XXX

Subcellular locations



Keywords



News



Forthcoming changes

Planned changes for UniProt

UniProt release 2015_11

The sense of a motion | Change of the cross-references to eggNOG

UniProt release 2015_10

The smell of the sea in UniProtKB | Cross-references to WBParaSite | Removal of the cross-references to CYGD | UniParc

[News archive](#)

Ensembl peptide - InterPro

Exons Alternating exons Alternating exons Residue overlap splice site

```
MTEDSDSISEEERSLFRPFTRESLQAI EARI ADEEAKQRELERKRAEGEDEDEGPQPDPT
LEQQVVPVPRMQGSFPPELASTPLEDI DSFYSNQRTFVVI SKGKDI FRFSATNALYVLDP
FNPIRRVAIYILVHPLFSFLIITTLVNCILMIMPTTPTVESTEVIFGTIYFESAVKVM
ARGFILQPFYLRDAMNWLDFVVI ALAYVTMGI DLGNLAALRTRFVRLAKLTVAIIVPLK
TIVGAVIESVKNLRDVIILTMSLSVFMALGLQIYMGVLTQKCIKEFPDLSGWNLTDES
WELFNSNDTNWFYSESGLDIPLCGNSSGAGQCDEGYICLQGYGKNPNYGYTSFDTFGWAF
SAFRLMTQDYWENLYQLVLR SAGPWHMLFFIV IIFLGSFYLVNLI LAIVMSYDELQKKA
EEEEAAEEEAALREAEAAAAA AAKLEAQAAAAAAAANPEIAKSPSDFSCHSYELFVQEK
GNDDNNEKMSIRSEGLEVSVEI TRTTAPTATAAGTAKARKVSAASLSLPGSPFNLRGGS
RGS HQFTIRNGRGRFVGPVGS DRKPLVLS TYLDAQEHL PYADDSNAVT PMSEENGAI IVP
VYYANLGRHSYSYTS HQSRI SYTSHGDL LGGMTKESRLRNR SARNTNHS I VPPNANNLS
VAETNHKGQRDFDLTQDC TDDAGKI KHNNDPF IEPAQTQTVVDMKDVMLVND I IEQAAGR
HSRASDHGVS VYYYFPTE DDEDEDGPTFKDKALEFLMKMIDIFCVWDCCVWLKFKQEVVAFI
VFDPFVELFITLCI VVNTLFMALDHDMDPMEKALKSGNYFFATATFAIEATMKLIAMSP
KYYFQEGWNI FDFI IVALS LLELGEVQGLSVLRSFRLLRVFKLAKSWPTLLNLSIMG
RTV GALGNLTFVLCII IIFIVAVMGQLFGKNYTDNVRDFPDHDLPRWNFTDFMHSFMIVF
RVLCGEWIESMWD CMLVGDVSCI PFFLATVVI GNLVNLFLALLSNFGSSLSAPTAD
NETNKIAEAFNRI SRFNSNWI KMNLANALKFVKNKLT SQIASVQPTGKGVCPICISSEHEN
EELTPDDI LADGLLKKGI KEHNQLEVAIGDGMEFTIHGDLKKNKAKKNQIMNNSKDDDT
ASIKSYGSHKNRPFKDESHKGS AETMEGEEKRDASKEDLGIDEELDDGEDEGDEGLDGE
L I IHAEEDEVIDS PADCCPDNCKYKFPVLAGDDDAPFWQGWGNLRLKTFQLIENKYFETA
VITMILLSS LALALEDVHLPQRPI LQDILY YMDRI FTVI FFLEMLIKNLALGFKVYFTNA
WCWLDFI IVMVSLIN FVASLCGAGGIQAFKTMRTL RALRPLRAMSRMQGMRVVVNLVQA
IPSIFNVLLVCLIFWLI FAIMGVQLFAGKYFKCV DKNK TLPHEI IIPDVNACKAENYSWE
NSPMNFDHV GKAYLCLCFQVATFKGWI QIMNDAI DSRDVS FVGKQPIRETNI YMYLYFVFF
IIFG SFFTLNLFIGVI I DNFNQK KKA GGSLEMFMTE DQKYYNAMKMKMGSKKPLKAI PR
PRWRPQAI VFEI VTNK KFDMI IMLFI GFNMLTMTLDHYKQSETF SAVLDYLNMFICIF S
SECLMKIFALRYHYFIEPNWLFDFVVI LLSILGLVLSDI IEKYFVSP TLLRVVRVAKVGR
VLRLVKGAKGIR TLLFALAMSLPALFNICLLLFVLMVFI FAIFGMSFFMHVKDKSGLDDVY
NFKTFGQSMI LLFQMST SAGWDGVL DGI INEEDCLP PDNDKGYPGNCGSSITIGIT YLLAY
LVISFLIVINMYIAVILENYSQATE DVQEG LTTDDYDMYIEIWQQFDPDGTQYVRYDQLS
DFLDVLEPPLQIHKPNRYKI ISMDIPI CRGDMMFVCDILDALTKDF FARKGNPIEETAEL
GEVQQRPEDEVGYE PVSS TLWRQREYCARLIQHAWKRYKQRHGGGT DASGDDLEI DACDN
GCGGGNGNENDDSGDGATGSGDNGSQHG GGSISGGGGT PGGGKSKGI IGSTQANIGIVDS
NISPKESPDSIGDPQGRQTAVLVESDGFVTKNGHRVVIHSRSPSITSRTADV
```

EMBL-EBI



InterPro
Protein sequence analysis & classification

Search InterPro...

Search

Examples: IPR020405, kinase, P51567, PF02932, GO:0007165

Home Search Release notes Download About InterPro Help Contact

InterPro: protein sequence analysis & classification

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. We combine protein signatures from a number of member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool. [Read more about InterPro](#)

Analyse your protein sequence

Search

Clear

Example protein sequence

v.54 InterPro 54.0
15th October 2015

Features include:

- An update to PANTHER (10.0).
- Integration of 2138 new methods from the CATH-Gene3D (2), Pfam (1319), SUPERFAMILY (5), ProDom (7) and PANTHER (805) databases.

[Download](#) | [Read more](#)

IDA
Domain architecture search
Search >>

Ensembl peptide - Interpro

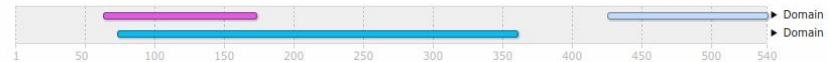
Exons Alternating exons Alternating exons Residue overlap splice site

```
MTEDSDSISEEERSLFRPFTRESLQAI EARI ADEEAKQRELERKRAEGEDEDEGPQPDPT
LEQGVVPPVRMQGSFPPELASTPLEDIDSFYSNQRTEVVVISKGDIFRFSATNALYVLDP
FNPIRRVAIYILVHPLFSFLIITTLVNCILMIMPPTTPTVESTEVIFGTIYFESAVKVM
ARGFILQPFPTYL RDAMNWLDFVVI ALAYVTMGI DLGNLALRTRFRVRLAKLTVAIIVPLK
TIVGAVIESVKNLRDVIILTMFSLSVFALMGLQIYMGVLTQKCIKEFPDLGDSWGNLTDES
WELFNSNDTNWFYSESGDIPLCGNSSGAGQCDEGYICLQGYGKNPNYGYTSFDTFGWAFI
SAFRMLTQDYWENLYQLVLR SAGPWHMLFFIV IIFLGSFYLVNLI LAIVAMS YDELQKKA
EEEEAEAEALREAEAEAAAAAKLEAQQA AAAANPEIAKSPSDFSCHSYELFVQGQEK
GNDDNKEKMSIRSEGLEVSVEI TRTTAPTATAAGTAKARKVSAASLSLPGSPFNLRGGS
RGS HQFTIIRNGRGRFVGVPGSDRKLPLVLSYLD AQEHL PYADDSNAVT PMSENGAI IVP
VYYANLGRHSYSYTS HQSRISYTS HGDLLGGMTEKSR LRNR SARNTNHSI VPPPNANLS
YAE TNHGQRDFDLTQDC TDDAGIKHNDNPFIEPAQTQTVVDMKDVMLNDIIEQAAGR
HSRASDHGVS VYVYFPTEDEDDDEGPTFKDKALEFLMKMIDIFCVWDC CWVWLKFGQEVAFI
VDFPVELFITLCI VVNTLFMALDHDMDPMEKALKSGNYFFATFAIEATMKLIAMSP
KYFYQEGWNI FDFIIVALSLLELGEVGGLSVLRFRLLRVFKLAKSWPTLNLLISIMG
RTV GALGNLTFVLCIIIF IFAVMGMQLFGKNYTDNVD RFPDHLPRWNFTDFMHSFMIVF
RVLCGEWIESMWDCLVGDVSCIPFFLATVVI GNLVVLNLF LALLSNFGSSLSAPTAD
NETNKIAEAFNRISRFSNWIKMNLANALKFVKNKLT SQIASVQPTGKGVCPISSEHGN
ELELTPDDI LADG LLKGIKEH NQLEVAIGDGM EFTIHGDLKNKAKKNKQIMNNSKDDDT
ASIKSYGSHKNRPFKDESHKGS AETMEGEEKRDASKEDLGIDE LDDGEDEGDEGLDDEL
IIHAEEDEVIDSDPADCCPDNCKYKFPVLAGDDAPFWQGWGNLRKLT FQLIENKYFETA
VITMILLSS LALALEDVHL PQRPI LQDILLYMDRI FTVIFLEMLIKNLALGFKVYFNA
WCWLDFIIVMVSLIN FVASLCGAGGIQAFKTMRTL RALRPLRAMSRMQGMRVVVNALVQA
IPSIFNVLLVCLIFWLIFAIMGVQLFAGKYFKCVDKNKTLPHEIIPD VNACKAENYSWE
NSPMMNFHDVHGKAYLCLFQVATFKGWIQIMND AIDSRDVSFVGKQPIRETNIYMYLYVFFF
IIFGSFFTLNLFIGVIIDNFNEQKKKAGGSEMFMTEDQKKYYNAMKMKMGSKKPLKAI PR
PRWRPQAVIFEIVTNK KFDMIIMLFI GFNMLTMTLDHYKQSETFSAVLDYLNMFICIF S
SECLMKIFALRYHYFIEPWNLFDFV VVILSLGLVLSDIIEKYFVSP TLLRVVRVAKVGR
VLRLVKGAKGIRTL L FALAMSLPALFNICLLLFVLMVIFAIFGMSFFMHVKDKSGLDDVY
NFKTFPGQSMILLFQMST SAGWDGVL DGIINEEDCLPDPNDKGYPGNGCSSTIGITYLLAY
LVISFLIVINMYIAVILENYSQATE DVQEGLTDDDYDMYIEIWQDFDPDGTQYVRYDQLS
DFLDVLEPPLQIHKPNRYKII SMDIPI CRGDMMFVCDIILDALTKDF FARKGNPIEETAEL
GEVQQRPEDEVGYE PVSS TLWRQRE EYCARLIQHAWKRYKQRHGGGT DASGDDLEI DACDN
GCGGGNGNENDDSGD GATGSGDNGS QHGSSISGGGT PGGGSKGIIGSTQANIGIVDS
NISPKESPDSIGDPQGRQTAVLVESDGFVTKNGHRVV IHSRSPSITSRTADV
```

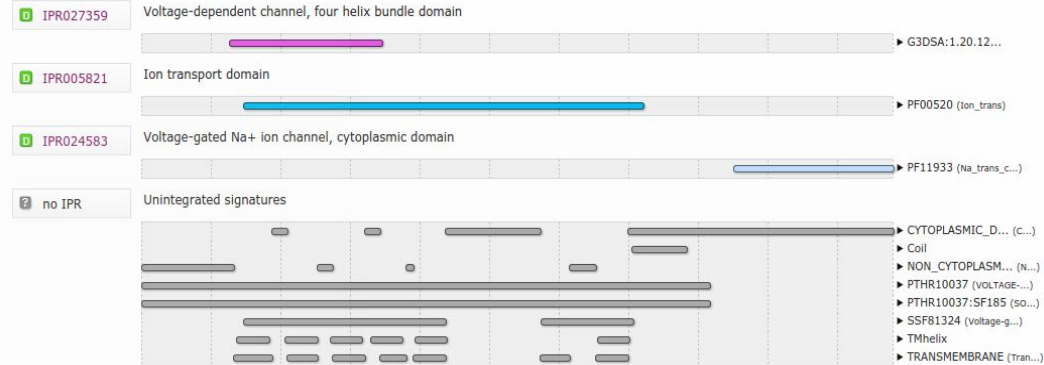
Protein family membership

None predicted.

Domains and repeats



Detailed signature matches



GO term prediction

Biological Process

- [GO:0006811](#) ion transport
- [GO:0055085](#) transmembrane transport

Molecular Function

- [GO:0005216](#) ion channel activity

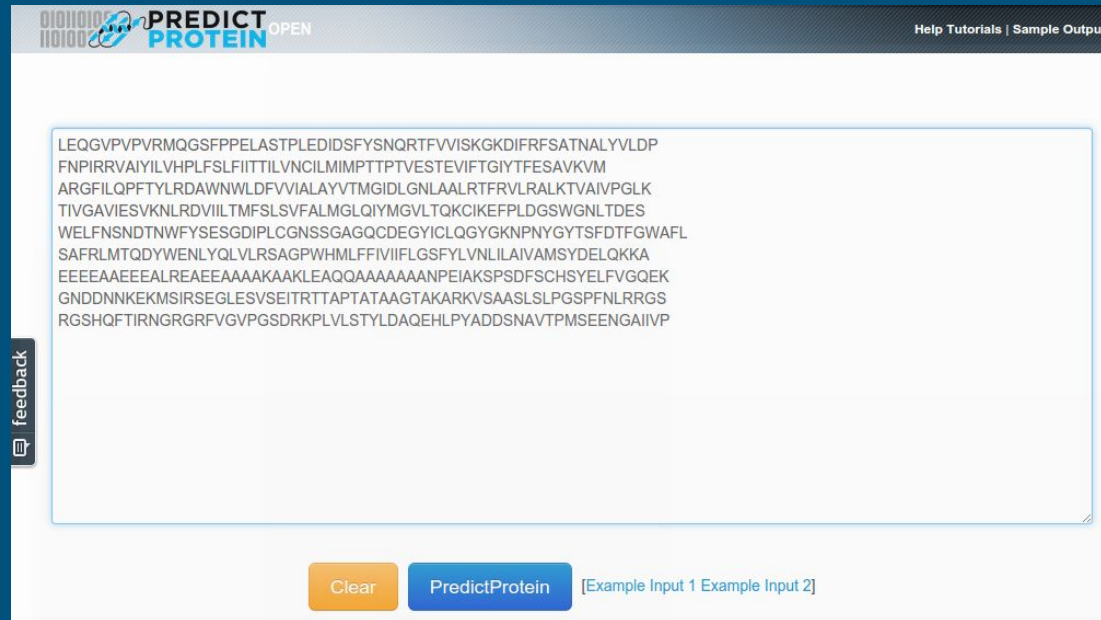
Cellular Component

- [GO:0016020](#) membrane

Ensembl - Genome Browser

Pfam	465	694	Domain of unknown function DUF3451	PF11933	IPR024583 [Display all genes with this domain]
PROSITE profiles	1834	1869	EF-hand domain	PS50222	IPR002048 [Display all genes with this domain]
Pfam	164	409	Ion transport domain	PF00520	IPR005821 [Display all genes with this domain]
Pfam	820	1004	Ion transport domain	PF00520	IPR005821 [Display all genes with this domain]
Pfam	1291	1516	Ion transport domain	PF00520	IPR005821 [Display all genes with this domain]
Pfam	1610	1816	Ion transport domain	PF00520	IPR005821 [Display all genes with this domain]
Pfam	1020	1265	Sodium ion transport-associated	PF06512	IPR010526 [Display all genes with this domain]
Prints	1799	1818	Voltage gated sodium channel, alpha subunit	PR00170	IPR001696 [Display all genes with this domain]
Prints	1834	1847	Voltage gated sodium channel, alpha subunit	PR00170	IPR001696 [Display all genes with this domain]
PANTHER	58	450	Voltage gated sodium channel, alpha-4 subunit	PTHR10037:SF193	IPR028826 [Display all genes with this domain]
PANTHER	477	491	Voltage gated sodium channel, alpha-4 subunit	PTHR10037:SF193	IPR028826 [Display all genes with this domain]

PredictProtein Open



The screenshot displays the PredictProtein Open web interface. At the top left, the logo features a stylized protein structure and the text "PREDICT PROTEIN OPEN". At the top right, there are links for "Help" and "Tutorials | Sample Output". The main area contains a large text input field with a protein sequence. Below the input field are two buttons: an orange "Clear" button and a blue "PredictProtein" button. To the right of the "PredictProtein" button is a small text label: "[Example Input 1 Example Input 2]". On the left side of the interface, there is a vertical "feedback" button with a speech bubble icon.

PredictProtein OPEN Help Tutorials | Sample Output

```
LEQGVVPVPRMQGSFPPELASTPLEDIDSFYSNQRTEFVVISKGGKIFRFSATNALYVLDP
FNPIRRVAIYLHPLFSLFIITLIVNCILMIMPTTPTVESTEVIPTGIYTFESAVKVM
ARGFILQPFTYLRDAWNWLDVVIALAYVTMGIDLGNLAALRFRVLRALKTVAIVPGLK
TIVGAVIESVKNLRDVIILTMFSLSVFALMGLQIYMGVLTQKCIKEFPLDGSWGNLTDES
WELFNNDTNWFYSESGDIPLCGNSSGAGQCDEGYICLQGYGKNPNYGYTSFDTFGWAFI
SAFRLMTQDYWENLYQLVRSAGPWHMLFFVIFLGSFYLVNLI LAIVAMSYDELQKKA
EEEEEEEEALREAAAAA KAAKLEAQAAAAAANPEIAKSPSDFSCHSYELFVGQEK
GNDDNNKEKMSIRSEGLSEVSEITRTTAPTATAAGTAKARKVSAASLSLPGSPFNLRGGS
RGSHQFTIRNGRGRFVGVPGSDRKPLVLSTYLD AQEHLPYADDSNAVTPMSEENGAIIVP
```

feedback

Clear PredictProtein [Example Input 1 Example Input 2]

[http://ppopen.informatik.tu-muenchen.de/visual_results?req_id=\\$1\\$04oxht0J\\$DgyLijzkqNhOOVVtxuHtR0#](http://ppopen.informatik.tu-muenchen.de/visual_results?req_id=$1$04oxht0J$DgyLijzkqNhOOVVtxuHtR0#)

PredictProtein Open

Dashboard >

STRUCTURE ANNOTATION

Secondary Structure and Solvent Accessibility >

Transmembrane Helices >

Protein Disorder and Flexibility >

Disulphide Bridges >

FUNCTION ANNOTATION

Effect of Point Mutations >

Gene Ontology Terms >

Subcellular Localization >

Binding Sites >

ADDITIONAL SERVICES

Literature Search >

HELP

Site Tutorial >

Molecular Function Ontology

#	GO ID	GO Term	Reliability (%)	
1	GO:0022838	substrate-specific channel activity	54	✓
2	GO:0022857	transmembrane transporter activity	54	✓
3	GO:0022803	passive transmembrane transporter activity	54	✓
4	GO:0005215	transporter activity	54	✓
5	GO:0005216	ion channel activity	54	✓
6	GO:0022891	substrate-specific transmembrane transporter activity	54	✓
7	GO:0022892	substrate-specific transporter activity	54	✓
8	GO:0015075	ion transmembrane transporter activity	54	✓
9	GO:0015267	channel activity	54	✓
10	GO:0022832	voltage-gated channel activity	51	✓

Navigation: Home, Previous, 1/9, Next, Home, 10

Cellular Component Ontology

#	GO ID	GO Term	Reliability (%)	
1	GO:0016020	membrane	60	✓
2	GO:0005886	plasma membrane	35	✓
3	GO:0044464	cell part	35	✓
4	GO:0005623	cell	35	✓
5	GO:0071944	cell periphery	35	✓
6	GO:0034706	sodium channel complex	33	✓
7	GO:0034702	ion channel complex	33	✓
8	GO:0034703	cation channel complex	33	✓
9	GO:1902495	transmembrane transporter complex	33	✓
10	GO:0044425	membrane part	33	✓

Navigation: Home, Previous, 1/8, Next, Home, 10

Biological Process Ontology

#	GO ID	GO Term	Reliability (%)	
1	GO:0006811	ion transport	57	✓
2	GO:0044699	single-organism process	57	✓
3	GO:0044765	single-organism transport	57	✓
4	GO:0006810	transport	46	✓
5	GO:0051234	establishment of localization	46	✓
6	GO:0051179	localization	46	✓
7	GO:0030001	metal ion transport	42	✓
8	GO:0006812	cation transport	42	✓
9	GO:0015672	monovalent inorganic cation transport	42	✓
10	GO:0006814	sodium ion transport	28	✓

Navigation: Home, Previous, 1/18, Next, Home, 10

PredictProtein Open

Dashboard >

STRUCTURE ANNOTATION

Secondary Structure and Solvent Accessibility >

Transmembrane Helices >

Protein Disorder and Flexibility >

Disulphide Bridg

FUNCTION ANNO

Effect of Point

Gene Ontology

Subcellular Loc

Binding Sites

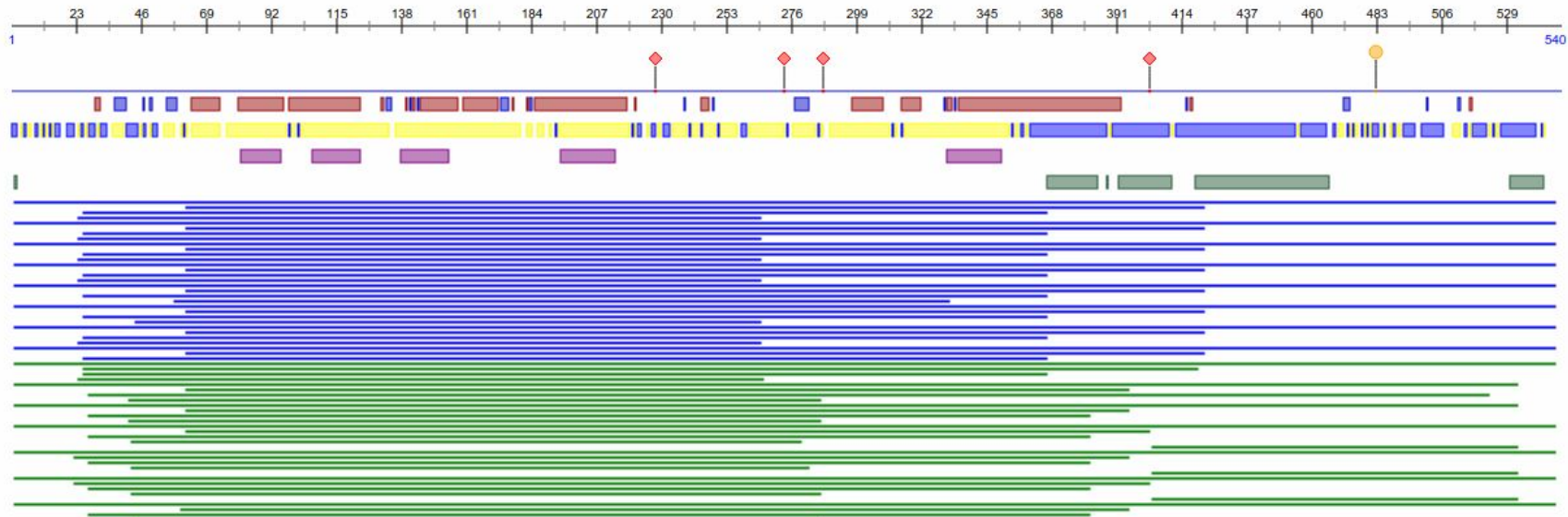
ADDITIONAL SER

Literature Sear


HELP

Site Tutorial

feedback



Gene Ontology and KEGG annotation



DAVID Bioinformatics Resources 6.7

National Institute of Allergy and Infectious Diseases (NIAID), NIH

[Home](#) [Start Analysis](#) [Shortcut to DAVID Tools](#) [Technical Center](#) [Downloads & APIs](#) [Term of Service](#) [Why DAVID?](#) [About Us](#)

Shortcut to DAVID Tools

- Functional Annotation**
Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and more
- Gene Functional Classification**
Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. [More](#)
- Gene ID Conversion**
Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. [More](#)
- Gene Name Batch Viewer**
Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. [More](#)

Recommending: A [paper published in Nature Protocols](#) describes step-by-step procedure to use DAVID!

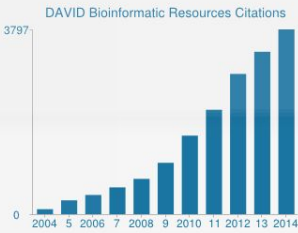
Welcome to DAVID 6.7

2003 - 2015

What's Important in DAVID?

- Current (v 6.7) [release note](#)
- [New requirement to cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supported](#)
- [Novel Classification Algorithms](#)
- [Pre-built Affymetrix and Illumina backgrounds](#)
- [User's customized gene background](#)
- [Enhanced calculating speed](#)

Statistics of DAVID



Year	Citations
2004	~100
2005	~200
2006	~300
2007	~400
2008	~500
2009	~700
2010	~1000
2011	~1500
2012	~2200
2013	~2800
2014	3797

- > 17,000 Citations
- Daily Usage: ~1200 gene lists/sublists from ~400 unique researchers.
- Total Usage: ~800,000 gene lists/sublists from >5,000 research institutes world-wide

Identify enriched biological themes, particularly GO terms

Discover enriched functional-related gene groups




Cluster redundant annotation terms

Visualize genes on BioCarta & KEGG pathway maps


Display related many-genes-to-many-terms on 2-D view.

Search for other functionally related genes not in the list

- List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs
- Redirect to related literatures
- Convert gene identifiers from one type to another.
- And more



Screen Shot 1 Screen Shot 2 Screen Shot 3



KEGG [Help](#)

[Japanese](#)

KEGG Home

[Release notes](#)
[Current statistics](#)
[Plea from KEGG](#)

KEGG Database

[KEGG overview](#)
[Searching KEGG](#)
[KEGG mapping](#)
[Color codes](#)

KEGG Objects

[Pathway maps](#)
[Brite hierarchies](#)

KEGG Software

[KegTools](#)
[KEGG API](#)
[KMGL](#)

KEGG FTP

[Subscription](#)

GenomeNet

[DBGET/LinkDB](#)

Feedback

[Kanehisa Labs](#)

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (See Release notes for new and updated features).

New article - KEGG as a reference resource for gene and protein annotation

- Main entry point to the KEGG web service**
 - KEGG2** [KEGG Table of Contents](#) [Update notes](#)
- Data-oriented entry points**
 - KEGG PATHWAY** [KEGG pathway maps](#) [[Pathway list](#)]
 - KEGG BRITE** [BRITE functional hierarchies](#) [[Brite list](#)]
 - KEGG MODULE** [KEGG modules](#) [[Module list](#) | [Statistics](#)]
 - KEGG ORTHOLOGY** [Orthology groups](#) [[KO system](#) | [Annotation](#)]
 - KEGG GENOME** [Genomes](#) [[KEGG organisms](#)]
 - KEGG GENES** [Genes and proteins](#) [[Release history](#)]
 - KEGG COMPOUND** [Small molecules](#) [[Compound classification](#)]
 - KEGG REACTION** [Biochemical reactions](#) [[Reaction modules](#)]
 - KEGG DISEASE** [Human diseases](#) [[Cancer](#) | [Pathogen](#)]
 - KEGG DRUG** [Drugs](#) [[ATC drug classification](#)]
 - KEGG MEDICUS** [Health information resource](#) [[Drug labels search](#)]
- Organism-specific entry points**
 - KEGG Organisms** Enter org code(s) [hsa](#) [hsa eco](#)
- Analysis tools**
 - KEGG Mapper** [KEGG PATHWAY/BRITE/MODULE mapping tools](#)
 - KEGG Atlas** [Navigation tool to explore KEGG global maps](#)
 - BlastKOALA** [Genome annotation and KEGG mapping](#)
 - GhostKOALA** [Metagenome annotation and KEGG mapping](#)
 - BLAST/FASTA** [Sequence similarity search](#)
 - SIMCOMP** [Chemical structure similarity search](#)

Galaxy Server @ IMBB

NGS experiment analysis

Flexible workflow designing

Genome annotation

RNA structure prediction

Very active community

70+ public servers available

Chipster Analysis Platform @ IMBB

Next Generation Sequencing

Genome browser

Alignment

Variants

RNA-seq: Assembly & differential expression

ChIP-seq: Peak detection motif scanning

Microarrays & Proteomics

Normalization

Filtering

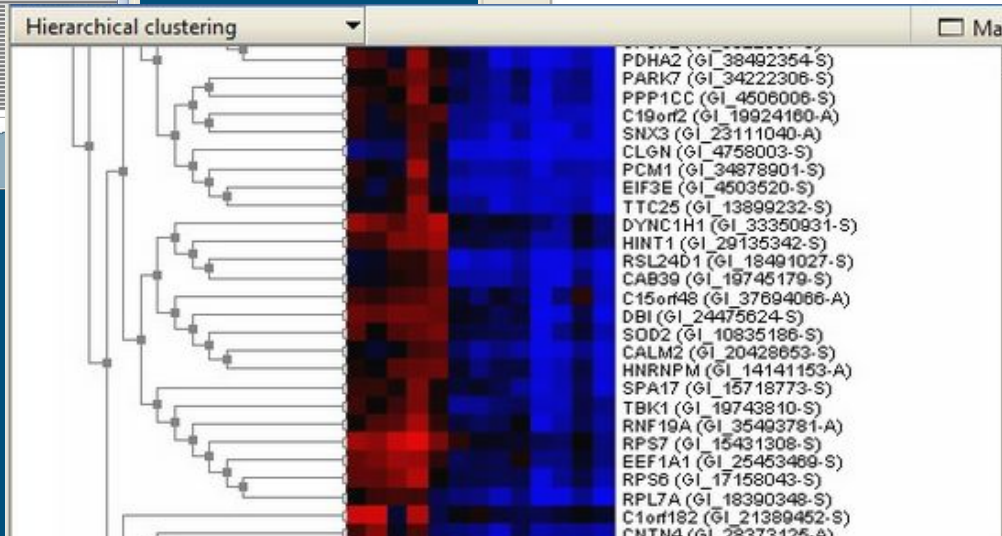
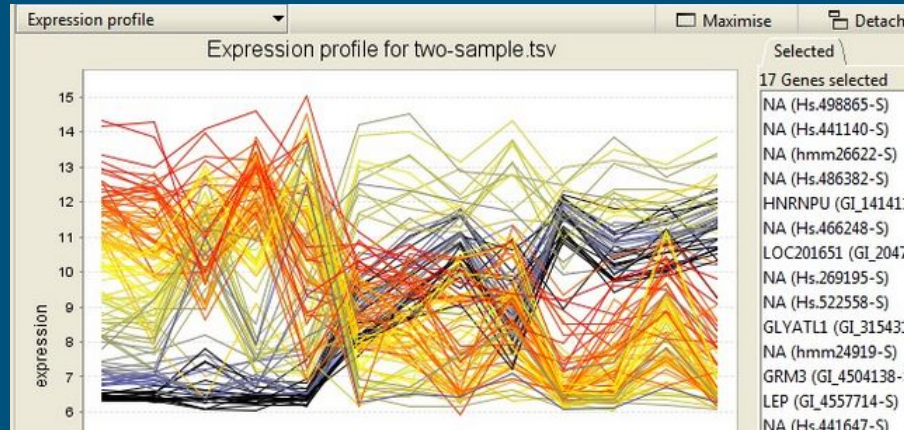
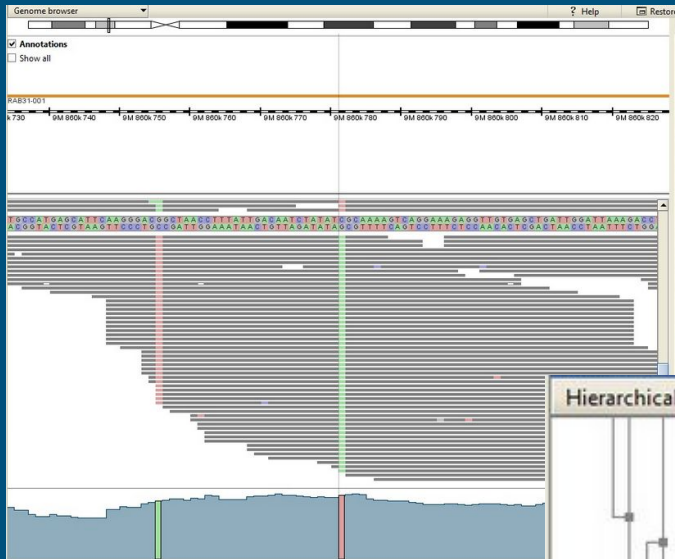
Plots (Venn, scattered, idiograms etc.)

Clustering

Pathway analysis (GO, KEGG, PFAM)



Chipster Analysis Platform



Thank you :-)

bioinfo@imbb.forth.gr

ENA - <http://www.ebi.ac.uk/ena>

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

InterPro - <http://www.ebi.ac.uk/interpro>

Predict Protein - <http://ppopen.informatik.tu-muenchen.de>
