



UNIVERSIDADE FEDERAL DA BAHIA  
Instituto Multidisciplinar em Saúde  
Campus Anísio Teixeira



# **Bancos de dados Moleculares**

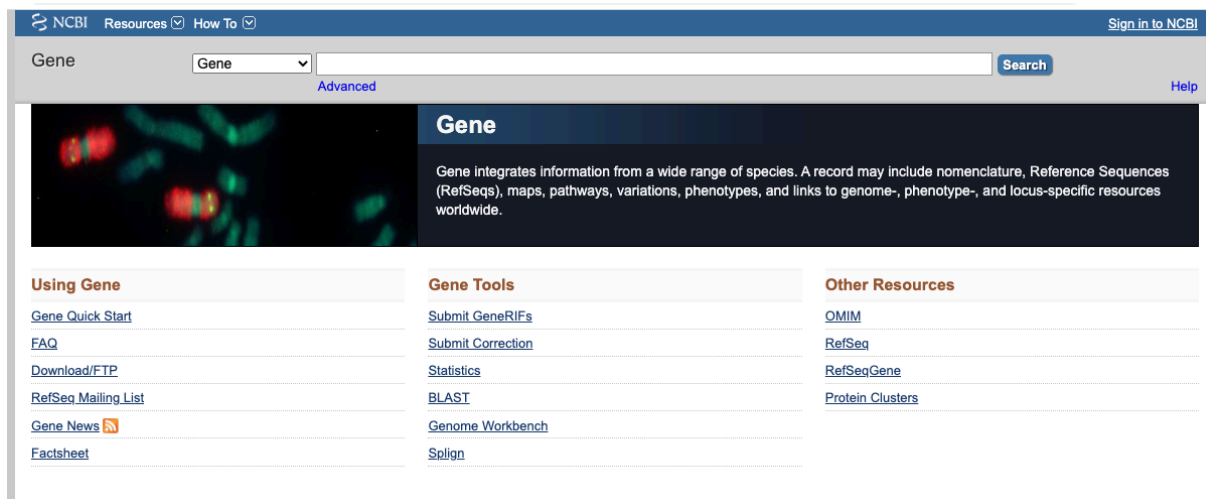
## **Bioinformática**

**Prof. Dr. Leandro Martins de Freitas**



Banco de dados: NCBI/Gene

Endereço: <https://www.ncbi.nlm.nih.gov/gene/>



A busca pode ser realizada pelo símbolo do gene. Usaremos como exemplo o gene CACNA1C.

Gene: Calcium voltage-gated channel subunit alpha1 C

Símbolo: CACNA1C

O banco NCBI/Gene encontrando um gene humana irá apresentar informações sobre esse gene logo no início do resultado da busca.



Abaixo será apresentado uma lista com os resultados encontrados para o símbolo pesquisado.



Search results

Items: 1 to 20 of 483

See also 8 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">CACNA1C</a> ID: 775	calcium voltage-gated channel subunit alpha1 C [ <i>Homo sapiens</i> (human)]	Chromosome 12, NC_000012.12 (1969552..2697950)	CACH2, CACN2, CACNL1A1, CCHL1A1, CaV1.2, LQT8, TS, TS. LQT8	114205
<input type="checkbox"/> <a href="#">Cacna1c</a> ID: 12288	calcium channel, voltage-dependent, L type, alpha 1C subunit [ <i>Mus musculus</i> (house mouse)]	Chromosome 6, NC_000072.7 (118564201..119174345, complement)	Cav1, Cav1.2, Cchl1, Cchl1a1, D930026N18Rik, MBC, MELC-CC	
<input type="checkbox"/> <a href="#">Cacna1c</a> ID: 24239	calcium voltage-gated channel subunit alpha1 C [ <i>Rattus norvegicus</i> (Norway rat)]	Chromosome 4, NC_051339.1 (151764138..152379454, complement)	RATIVS302	
<input type="checkbox"/> <a href="#">cacna1c</a> ID: 170581	calcium channel, voltage-dependent, L type, alpha 1C subunit [ <i>Danio rerio</i> (zebrafish)]	Chromosome 4, NC_007115.7 (20309073..20480289)	C-LTC, C-LTCC, ZfCav1, isl	

Essa lista possui o gene pesquisado encontrado em vários organismos. Na lista podemos identificar a descrição do gene e juntamente a espécie que esse gene é encontrado (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, e *Danio rerio*). Também é apresentado a localização do gene no genoma de cada espécie, Aliases mostra outros símbolos já usados para identificar o gene, e o MIM (Mendelian Inheritance in Man) que apresenta um link para o banco de genes de herança mendeliana em humanos.

Summary

CACNA1C calcium voltage-gated channel subunit alpha1 C [ *Homo sapiens* (human) ]

Download Datasets

Gene ID: 775, updated on 10-Apr-2021

Summary

Official Symbol

CACNA1C provided by HGNC

Official Full Name

calcium voltage-gated channel subunit alpha1 C provided by HGNC

Primary source

HGNC:HGNC:1390

See related

Ensembl:ENSG00000151067 MIM:114205

Gene type

protein coding

RefSeq status

REVIEWED

Organism

Homo sapiens

Lineage

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as

TS; LQT8; CACH2; CACN2; CaV1.2; CCHL1A1; CACNL1A1; TS. LQT8

Summary

This gene encodes an alpha-1 subunit of a voltage-dependent calcium channel. Calcium channels mediate the influx of calcium ions into the cell upon membrane polarization. The alpha-1 subunit consists of 24 transmembrane segments and forms the pore through which ions pass into the cell. The calcium channel consists of a complex of alpha-1, alpha-2/delta, beta, and gamma subunits in a 1:1:1:1 ratio. There are multiple isoforms of each of these proteins, either encoded by different genes or the result of alternative splicing of transcripts. The protein encoded by this gene binds to and is inhibited by dihydropyridine. Alternative splicing results in many transcript variants encoding different proteins. Some of the predicted proteins may not produce functional ion channel subunits. [provided by RefSeq, Oct 2012]

Expression

Broad expression in endometrium (RPKM 5.2), heart (RPKM 4.5) and 19 other tissues See more

Orthologs

mouse all

NEW

Try the new Gene table

Try the new Transcript table

Genomic context



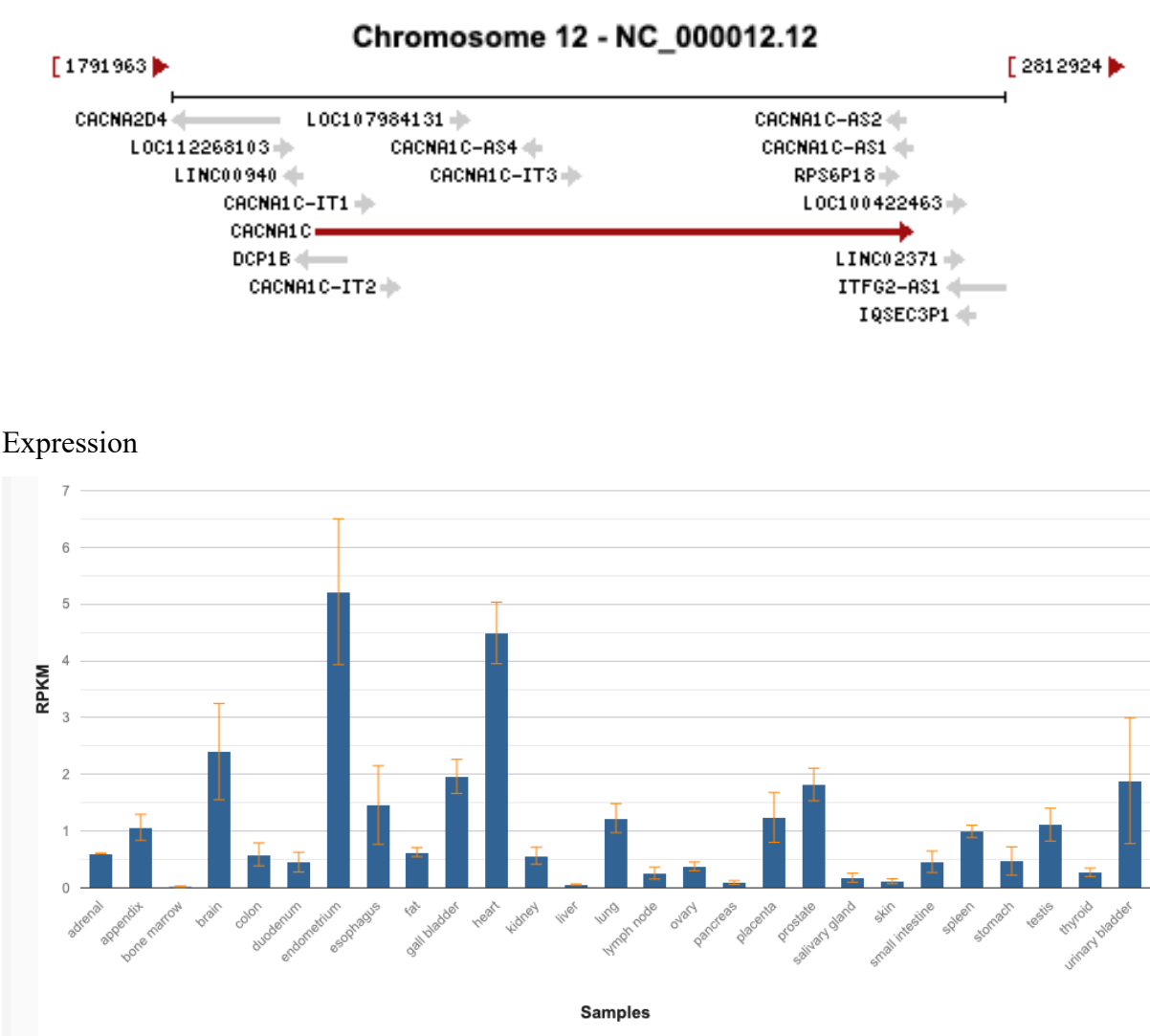
Informação sobre localização cromossômica e número de éxons que o gene possui

Location: 12p13.33

Exon count: 56

Localização inicial e final no cromossomo Identificador (início..fim):

NC\_000012.12 (1969552..2697950)



General gene information

Gene Ontology (GO)

Molecular function

Biological process

Cellular component

NCBI Reference Sequences (RefSeq)



Genomic

1. NG\_008801.2 RefSeqGene

Range	87465..732164
Download	<a href="#">GenBank</a> , <a href="#">FASTA</a> , <a href="#">Sequence Viewer (Graphics)</a> , <a href="#">LRG 334</a>

mRNA and Protein(s)

1. NM\_000719.7 → NP\_000710.5 voltage-dependent L-type calcium channel subunit alpha-1C isoform 18

[See identical proteins and their annotated locations for NP\\_000710.5](#)

Status: REVIEWED

Description	Transcript Variant: This variant (18), also referred to as HLCC70, lacks three alternate in-frame exons, compared to variant 1, resulting in a shorter protein (isoform 18), compared to isoform 1.	
Source sequence(s)	<a href="#">AC005342</a> , <a href="#">AC007618</a> , <a href="#">Z34810</a>	
Consensus CDS	<a href="#">CCDS44794.1</a>	
UniProtKB/Swiss-Prot	<a href="#">Q13936</a>	
Related	<a href="#">ENSP00000382563.1</a> , <a href="#">ENST00000399655.6</a>	
Conserved Domains (5) <a href="#">summary</a>		
	<a href="#">pfam00520</a> Location:1236 → 1487	Ion_trans; Ion transport protein
	<a href="#">pfam08763</a> Location:1560 → 1631	Ca_chan_IQ; Voltage gated calcium channel IQ domain
	<a href="#">pfam16885</a> Location:1654 → 2086	CAC1F_C; Voltage-gated calcium channel subunit alpha, C-term
	<a href="#">pfam16905</a> Location:1495 → 1558	GPHH; Voltage-dependent L-type calcium channel, IQ-associated
	<a href="#">cl01503</a> Location:202 → 291	TrbL; TrbL/VirB6 plasmid conjugal transfer protein

CACNA1C, CHD8, FMR1, MECP2, NRXN1, PTEN SCN1/2A, SHANK2/3, TSC1/2, and UBE3A

O banco de dados NCBI/Nucleotide armazena uma coleção de sequências completas e parciais de DNA. As sequências de DNA podem ser genes ou regiões intergênicas, completas, parciais, RNA (CDS). Dessa forma o banco é bem redundante e não curado.

A busca pode ser feita por diversos termos como nome do gene, símbolo, identificador da entrada no banco de dados (ID), organismo. A busca simples irá retornar todas as sequências onde o termo pesquisado foi encontrado.



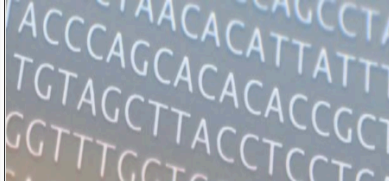
Nucleotide

Nucleotide

Advanced

Search

Help



### Nucleotide

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

Items: 1 to 20 of 4018

<< First < Prev Page 1 of 201 Next > Last >>

- ☐ [Cryptocotyle lingua voltage-dependent L-type calcium channel subunit alpha-1C \(CACNA1C\) mRNA, complete cds](#)
1. [alpha-1C \(CACNA1C\) mRNA, complete cds](#)  
8,497 bp linear mRNA  
Accession: MW361114.1 GI: 1964739067  
[Protein](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Symbiodinium sp. CCMP2456 genome assembly, contig: Slin CCMP2456.scaffold62, whole genome shotgun sequence](#)
2. [Slin CCMP2456.scaffold62, whole genome shotgun sequence](#)  
254,779 bp linear DNA  
Accession: CAJNDV010033552.1 GI: 1990883135  
[BioProject](#) [BioSample](#) [Protein](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

Vamos aplicar filtro de pesquisa para retornar somente o gene pesquisado para um grupo taxonômico. Vamos pesquisar o gene CACNA1C para a espécie *Homo sapiens*. Para isso basta usar o filtro ao lado direito da página. Nessa lista está os organismos com maiores resultados para o termo pesquisado. Se o seu organismo de interesse não está na lista você deve usar uma busca avançada (mais abaixo).



Filters: [Manage Filters](#)

## Results by taxon

### Top Organisms [\[Tree\]](#)

Homo sapiens (106)  
Canis lupus familiaris (83)  
Gallus gallus (80)  
Mus musculus (73)  
Cricetulus griseus (65)  
All other taxa (3611)

[More...](#)

Termo de busca gerado pela página: (CACNA1C) AND "Homo sapiens"[porgn: \_\_txid9606]

Resultado: 106 entradas no banco de dados NCBI/Nucleotide

Essa pesquisa retorna genes da espécie Homo sapiens. Entretanto, a busca não é específica e retorna outros genes onde o termo CACNA1C também é encontrado. A busca irá retornar qualquer entrada que contenha o termo pesquisado.

A busca também retorna o gene UBE2R2 antisense RNA 1 (UBE2R2-AS1) porque na descrição do gene está o título do arquivo “Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder” onde aparece o nosso gene pesquisado.

Podemos melhorar essa busca usando pesquisa mais específica usando a opção avançada.

Nucleotide

Nucleotide

CACNA1C

Search

[Create alert](#) [Advanced](#) [Help](#)

Species

Animals (3,843)

Plants (2)

Protists (163)

Customize ...

Molecule types

Summary

20 per page

Sort by Default order

Send to:

Filters: [Manage Filters](#)

Results by taxon

Top Organisms [\[Tree\]](#)

Homo sapiens (106)

Canis lupus familiaris (83)

Gallus gallus (80)

GENE

Was this helpful?

👍

👎

Pesquisa no banco de dados NCBI/Nucleotide usando a opção avançada.





#### Nucleotide Advanced Search Builder

(CACNA1C[Gene Name]) AND Homo sapiens[Organism]

[Edit](#)

[Clear](#)

##### Builder

Gene Name

CACNA1C

[Show index list](#)

AND

Organism

Homo sapiens

[Show index list](#)

AND

All Fields

[Show index list](#)

[Search](#) or [Add to history](#)

Construindo uma pesquisa avançada no NCBI/Nucleotide.

Vamos fazer a pesquisa pelo termo CACNA1C usando a opção: **Gene Name**

Vamos adicionar o grupo taxonômico Homo sapiens: **Organism**

Todo termo adicionado será acrescentado “AND” para fazer uma interseção do resultado. Nesse caso vamos fazer a busca de todas as entradas que tenham no **Nome do Gene** CACNA1C e ao mesmo tempo sejam do **Organismo** *Homo Sapiens*.

Após ter inserido os termos de busca nos campos clique no botão *Search*.

Termo de busca construído: (CACNA1C[Gene Name]) AND Homo sapiens[Organism]

Resultado: 68 entradas no banco de dados NCBI/Nucleotide.

As entradas possuem links para mostrar a sequência de nucleotídeos no formato GenBank, FASTA ou *Graphics*. O formato GenBank possui mais informações sobre a entrada, enquanto o formato FASTA possui somente uma linha de identificação simples e a sequência. Além disso existem também links para os bancos de proteína (*Protein*), informações sobre artigos onde a entrada foi encontrada (*PubMed*), informações taxonômicas da espécie que possui aquela sequência (*Taxonomy*).

Para baixar as sequências desejadas você deve selecionar as sequências e ir até a opção “enviar para” (*Send to*). Se desejar baixar todas as sequências vá direto para a “enviar para” (*Send to*) sem selecionar nenhuma sequência. As sequências podem ser baixadas para seu computador em diversos formatos. Sugiro baixar no formato FASTA que a maioria dos programas que analisam sequências aceita.



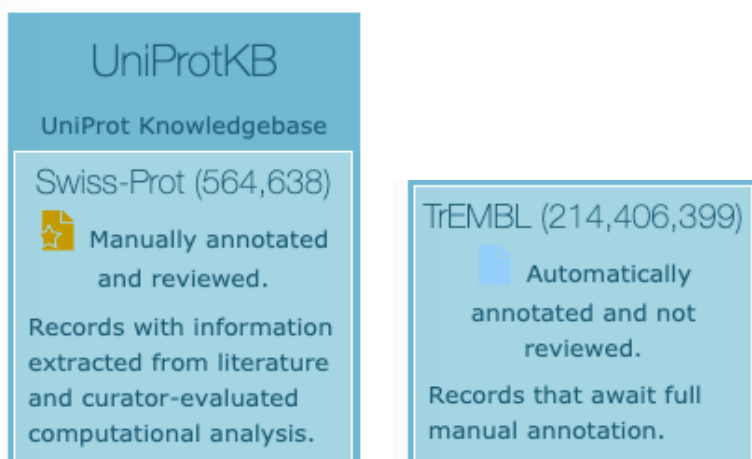
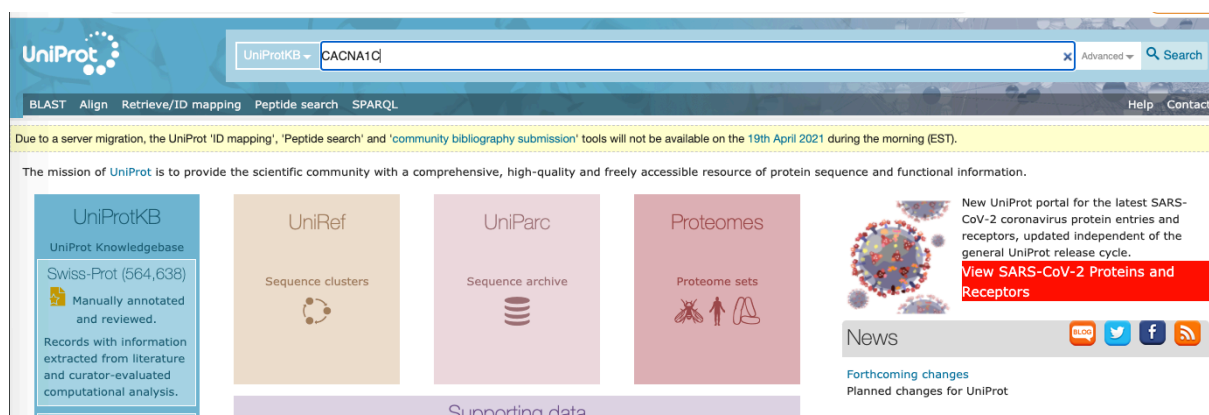


UNIVERSIDADE FEDERAL DA BAHIA  
Instituto Multidisciplinar em Saúde  
Campus Anísio Teixeira



Banco de dados: UniProt

Endereço: <https://www.uniprot.org>



Resultado da Busca:

Total de entradas:

Reviewed (Swiss-Prot): 133 entradas

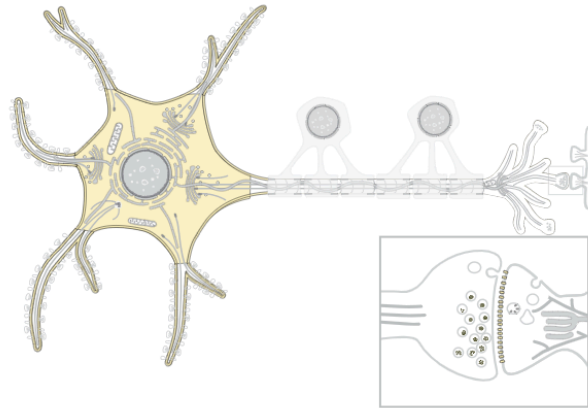
Unreviewed (TrEMBL): 2128 entradas

As sequências do Swiss-Prot foram revisadas e as informações são confiáveis. As informações do TrEMBL não foram revisadas e devem ser usadas com cuidado para não inferir novas informações falsas.



Subcellular location<sup>1</sup>

UniProt annotation    GO - Cellular component



- Plasma membrane
- Cell membrane <sup>i</sup> 20 Publications ; Multi-pass membrane protein <sup>i</sup> Curated
  - sarcolemma <sup>i</sup> By similarity ; Multi-pass membrane protein <sup>i</sup> Curated
  - postsynaptic density membrane <sup>i</sup> By similarity
  - T-tubule <sup>i</sup> By similarity
- Other locations
- Perikaryon <sup>i</sup> By similarity
  - dendrite <sup>i</sup> By similarity
- Note: Colocalizes with ryanodine receptors in distinct clusters at the junctional membrane, where the sarcolemma and the sarcoplasmic reticulum are in close contact. The interaction between RRAD and CACNB2 promotes the expression of CACNA1C at the cell membrane. By similarity

Automatic annotation  
Manual annotation  
Source: SwissBioPics

Localização da proteína na célula.

Filter by<sup>1</sup>

Reviewed (133)

Swiss-Prot

Unreviewed (2,128)

TrEMBL

Popular organisms

Mouse (62)

Human (44)

Rat (27)

Bovine (16)

Zebrafish (8)

Other organisms

Go

BLAST

Align

Download

Add to basket

Columns

1 to 25 of 2,261

Show 25

<div><div><div></div></div></div>	Entry	Entry name	<div><div><div></div></div></div>	Protein names	Gene names	Organism	Length	
<div><div><div></div></div></div>	<div><div></div><div>Q13936</div></div>	CAC1C_HUMAN	<div><div><div></div></div></div>	<div><div><div></div></div></div> <div>Voltage-dependent L-type calcium ch...</div>	CACNA1C CAC2, CACN2, CACNL1A1, CCHL1A1	Homo sapiens (Human)	2,221	
<div><div><div></div></div></div>	<div><div></div><div>Q01815</div></div>	CAC1C_MOUSE	<div><div><div></div></div></div>	<div><div><div></div></div></div> <div>Voltage-dependent L-type calcium ch...</div>	Cacna1c Cac2, Cacn2, Cacnl1a1, Cchl1a1	Mus musculus (Mouse)	2,139	
<div><div><div></div></div></div>	<div><div></div><div>P22002</div></div>	CAC1C_RAT	<div><div><div></div></div></div>	<div><div><div></div></div></div> <div>Voltage-dependent L-type calcium ch...</div>	Cacna1c Cac2, Cacn2, Cacnl1a1, Cchl1a1	Rattus norvegicus (Rat)	2,169	
<div><div><div></div></div></div>	<div><div></div><div>P15381</div></div>	CAC1C_RABIT	<div><div><div></div></div></div>	<div><div><div></div></div></div> <div>Voltage-dependent L-type calcium ch...</div>	CACNA1C CAC2, CACN2, CACNL1A1, CCHL1A1	Oryctolagus cuniculus (Rabbit)	2,171	
<div><div><div></div></div></div>	<div><div></div><div>O35505</div></div>	CAC1C_CAVPO	<div><div><div></div></div></div>	<div><div><div></div></div></div> <div>Voltage-dependent L-type calcium ch...</div>	CACNA1C CAC2, CACN2, CACNL1A1, CCHL1A1	Cavia porcellus (Guinea pig)	2,169	
<div><div><div></div></div></div>	<div><div></div><div>O73707</div></div>	CAC1C_CHICK	<div><div><div></div></div></div>	<div><div><div></div></div></div> <div>Voltage-dependent L-type calcium ch...</div>	CACNA1C	Gallus gallus (Chicken)	177	
<div><div><div></div></div></div>	<div><div></div><div>Q90ZA7</div></div>	Q90ZA7_DANRE	<div><div><div></div></div></div>	<div><div><div></div></div></div> <div>Voltage-dependent L-type calcium ch...</div>	cacna1c sl	Danio rerio (Zebrafish) (Brachydanio rerio)	2,168	
<div><div><div></div></div></div>	<div><div></div><div>Q5TZF1</div></div>	Q5TZF1_DANRE	<div><div><div></div></div></div>	<div><div><div></div></div></div> <div>Voltage-dependent L-type calcium ch...</div>	cacna1c	Danio rerio (Zebrafish) (Brachydanio rerio)	2,196	



## Prática

### Banco de dados Gene/NCBI

1. Entrar no banco de dados

<https://www.ncbi.nlm.nih.gov/gene>

2. Realizando a busca: Buscar pelo gene com símbolo PTEN

3. Resultado da Busca, primeira parte

4. Resultado da Busca, segunda parte


Clicar no símbolo do gene PTEN de humanos (*Homo sapiens*)



Search results

Items: 1 to 20 of 8335

See also 87 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases	MIM
 <a href="#">PTEN</a> ID: 5728	phosphatase and tensin homolog [Homo sapiens (human)]	Chromosome 10, NC_000010.11 (87863625..87971930)	10q23del, BZS, CWS1, DEC, GLM2, MHAM, MMAC11, PTENbeta, TEP1, PTEN	601728
<input type="checkbox"/> <a href="#">Pten</a> ID: 19211	phosphatase and tensin homolog [Mus musculus (house mouse)]	Chromosome 19, NC_000085.7 (32734977..32803560)	23100350O07Rik, A130070J02Rik, B430203M17Rik, MMAC1, PTENbeta, TEP1	
<input type="checkbox"/> <a href="#">Pten</a> ID: 50557	phosphatase and tensin homolog [Rattus norvegicus (Norway rat)]	Chromosome 1, NC_086019.1 (240043707..240110330)	MMAC1, Mmac, TEP1	

5. Informações sobre o gene PTEN de humanos

Gene

Gene

Advanced

Full Report

Send to:

**PTEN phosphatase and tensin homolog [ Homo sapiens (human) ]**

Download Datasets

Gene ID: 5728, updated on 31-Mar-2024

Summary

**Official Symbol** PTEN provided by HGNC  
**Official Full Name** phosphatase and tensin homolog provided by HGNC  
**Primary source** HGNC:HGNC:9588  
**See related** Ensembl:ENSG00000171862 MIM:601728; AllianceGenome:HGNC:9588  
**Gene type** protein coding  
**RefSeq status** REVIEWED  
**Organism** Homo sapiens  
**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo  
**Also known as** BZS; DEC; CWS1; GLM2; MHAM; TEP1; MMAC1; PTEN1; 10q23del; PTENbeta

Exercícios

Buscar as seguintes informações

Gene ID (Entrez ID):

Símbolo oficial (Official Symbol):

Nome oficial (Official Full Name):

Símbolos anteriores (Also known as):

Descreva de forma sucinta a função da proteína PTEN:




Qual cromossomo está localizada o gene:

Localização no cromossomo, bandas e sub-bandas:

Posição de início no cromossomo (versão GRCh38.p14):

Posição de final no cromossomo (versão GRCh38.p14)

Número de exons do gene:

Quais são os três tecidos que mais expressão esse gene (projeto HPA - tissue-specificity of all protein-coding genes):

- 1.
- 2.
- 3.

Existem condições associadas a esse gene? Liste as 3 primeiras.

- 1.
- 2.
- 3.

Quais os três primeiros processos que esse gene está associado (Gene Ontology - Process)

- 1.
- 2.
- 3.

RefSeq Genomic ID:

Quantas isoformas esse gene apresenta?

Qual o RefSeq ID da primeira isoforma

mRNA:

Proteína:

Recuperar a sequência fasta da isoforma 1 (CDS do gene)

## Banco de dados UNIPROT

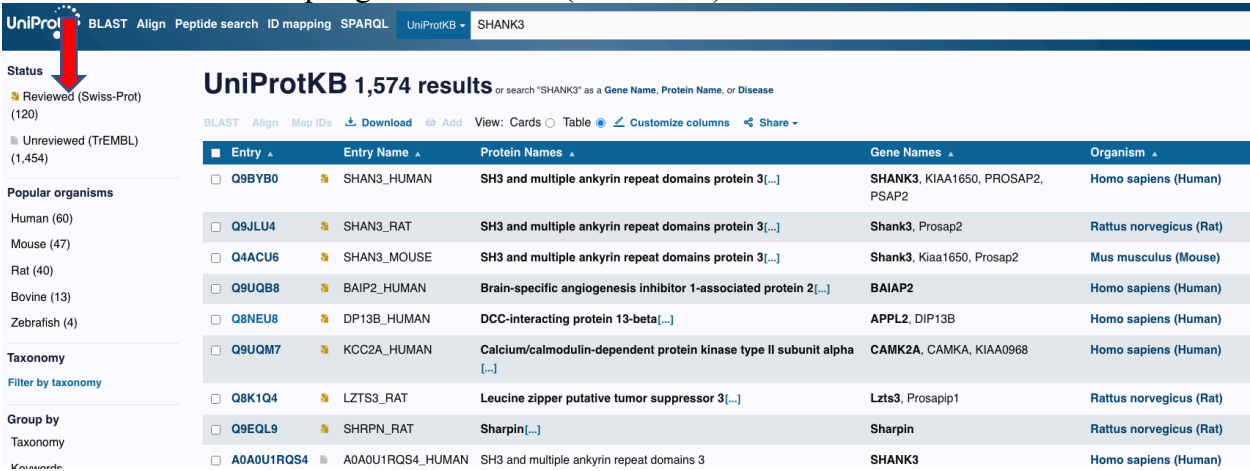
1. Entrar no banco de dados do UniProt

<https://www.uniprot.org/>



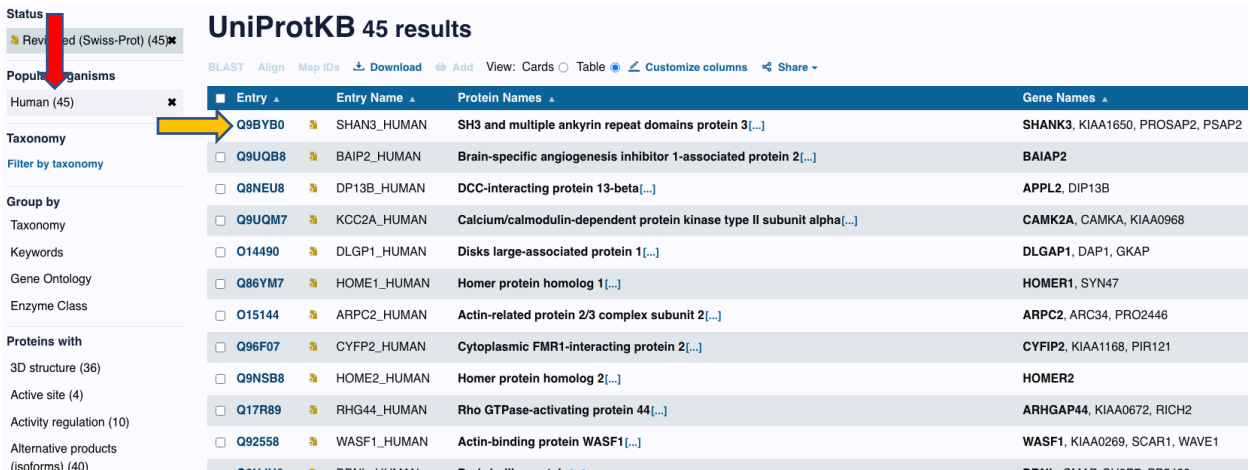
2. Fazer a busca pelo gene SHANK3

3. Refinar a busca por gene Revisados (Swiss-Prot)



4. Refinar a busca para organismo humano (Human - *Homo sapiens*) indicado pela seta vermelha

5. Vamos usar a proteína (realmente a proteína SHANK3) indicada pela seta amarela vermelha



## Exercícios

Qual o número de identificação do UniProt para essa sequência:

Quantos aminoácidos tem essa proteína:

Qual a evidência para a existência dessa proteína (Protein existence):

Símbolo oficial:



UNIVERSIDADE FEDERAL DA BAHIA  
Instituto Multidisciplinar em Saúde  
Campus Anísio Teixeira



Símbolos anteriores (Synonyms):

Descreva de forma sucinta a função da proteína SHANK3:


Quais são os três tecidos que mais expressão essa proteína:

- 1.
- 2.
- 3.

Existem condições associadas a essa proteína? Liste as 3 primeiras.

- 1.
- 2.
- 3.

Quais os três primeiros processos que essa proteína está associada (Gene Ontology – Biological Process)

- 1.
- 2.
- 3.

Localização sub-cellular:

Apresente a imagem da localização sub-cellular

Quais as duas condições associadas a essa proteína?

- 1.
- 2.

Quais tecidos expressam essa proteína (Expression Tissue specificity)

Quantas estruturas secundárias tipo hélice essa proteína apresenta:

	Início	Fim
1.		
2.		
3.		

\*adicionar mais linhas na tabela caso necessário