

# Biological Databases

Dr. Upeskha Ganegoda

# Other biological databases

- Swiss-prot (Uniprot)
- Protein databases
- Disease databases
- lncRNA databases

# Swiss-prot (Uniprot)

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 (551,705)**  
Manually annotated and reviewed.  
**TrEMBL (65,378,749)**  
Automatically annotated and not reviewed.

**UniRef**  
Sequence clusters

**UniParc**  
Sequence archive

**Proteomes**

**Supporting data**

Literature citations, Cross-ref. databases, Taxonomy, Diseases, Subcellular locations, Keywords

**News**

Forthcoming changes, Planned changes for UniProt, UniProt release 2016\_07, UniProt release 2016\_06

**Protein spotlight**

**Of Plastic And Men**, July 2016. Nature has extraordinary resources. Here we are trashing her land, sea and atmosphere - and have been for over a century now - with all sorts of chemistry she didn't ask for and which, sooner or later, will prove to be harmful to those who are putting it there. Despite this, sometimes she manages to find ways of twisting something bad into something good. Polyethylene terephthalate, also known as PET, is one...

## Getting started



### Text search

Our basic text search allows you to search all the resources available

### BLAST

Find regions of similarity between your sequences

### Sequence alignments

Align two or more protein sequences using the Clustal Omega program

### Retrieve/ID mapping

This tool merges the "Retrieve" and "ID Mapping" tools

## UniProt data

### Download latest release

Get the UniProt data

### Statistics

View Swiss-Prot and TrEMBL statistics

### How to cite us

The UniProt Consortium

### Submit your data

Submit your sequences and annotation updates

### SPARQL

Query UniProt data using a SQL like graph query language

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

# ID mapping

The screenshot shows the UniProt homepage with a blue header. The header includes the UniProt logo, a dropdown menu set to "UniProtKB", an "Advanced" search link, a search bar with a magnifying glass icon, and "Help Contact" links. Below the header, a navigation bar offers "BLAST", "Align", and "Retrieve/ID mapping". The main content area is titled "Retrieve/ID mapping" in orange. A section titled "How to use this tool" contains instructions and examples for entering identifiers. To the right, a box provides step-by-step instructions for using the tool.

## How to use this tool

Enter or upload a list of identifiers to do one of the following:

Retrieve the corresponding UniProt entries to download them or work with them on this website.

Convert identifiers which are of a different type to UniProt identifiers or vice versa and download the identifier lists.

1. Enter identifiers, separated by a space or a new line, into the form field, for example:

P31946 P62258

ALBU\_HUMAN

EFTU\_ECOLI

2. If you need to convert to another identifier type, select the source and target type from the dropdown menus.

3. Click the Go button.

[? Help](#) [Help video](#) [Other tutorials and videos](#) [Downloads](#)

## 1. Provide your identifiers

7448960

OR upload your own file:  No file selected.

Run in a new window.

## 2. Select options

From

To

GenID (Entrez Gene)

UniProtKB

Go

# UniProtKB results

[? About](#) [Upload lists](#) [Basket](#)

1 out of 1 Entrez Gene (GeneID) identifier was successfully mapped to 1 UniProtKB ID in the table below.

## Filter by

Unreviewed (1)  
TrEMBL

**Popular organisms**  
THAPS (1)

## View by

Taxonomy

Keywords

Gene Ontology

Enzyme class

Pathway

## UniRef

Your results in sequence  
clusters with identity of:  
100%, 90% or 50%

## Demo

Help video

	BLAST	Align	Download	Add to basket	Columns	>	1 to 1 of 1		Show 25
	Your list:...13EFF2	Isoform map:...13EFF2	Entry	Entry name		Protein names	Gene names	Organism	Length
	<input type="checkbox"/> 7448960		B5YME7	B5YME7_THAPS		Breast cancer 2 early onset	BRAC2 THAPS_263089	Thalassiosira pseudonana (Marine diatom) (Cyclotella nana)	323

1 to 1 of 1 Show 25

## Display

[BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#)

[Feedback](#) [Help video](#) [Other tutorials and videos](#)

### Entry

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Cross-references

Publications

Entry information

Miscellaneous

Similar proteins

▲ Top

Protein | Submitted name: **Breast cancer 2 early onset**

Gene | **BRAC2**

Organism | *Thalassiosira pseudonana* (Marine diatom) (*Cyclotella nana*)

Status |  Unreviewed - Annotation score: ⓘ 00000 - Protein predicted ⓘ

## Function

### GO - Molecular function

- single-stranded DNA binding 

### GO - Biological process

- double-strand break repair via homologous recombination 
- regulation of transcription, DNA-templated 

Complete GO annotation...

## Names & Taxonomy

Protein names ⓘ

Submitted name:

Breast cancer 2 early onset 

Gene names ⓘ

Name: BRAC2 

ORF Names: THAPS\_263089 

Organism ⓘ

Thalassiosira pseudonana (Marine diatom) (*Cyclotella nana*) 

Taxonomic identifier ⓘ

35128 [NCBI]

Taxonomic lineage ⓘ

Eukaryota > Stramenopiles > Bacillariophyta > Coscinodiscophyceae > Thalassiosiophycidae > Thalassiosirales > Thalassiosiraceae > Thalassiosira 

Proteomes ⓘ

UP000001449 Component ⓘ: Chromosome 7

## Display

### Entry

Feature viewer

Feature table

None

- Function
- Names & Taxonomy
- Subcellular location

Pathology & Biotech

PTM / Processing

Expression

- Interaction

Structure

- Family & Domains

- Sequence

Cross-references

Publications

Entry information

Miscellaneous

Similar proteins

▲ Top

### Subcellular location<sup>i</sup>

#### GO - Cellular component<sup>i</sup>

■ nucleus 

Complete GO annotation...

### Interaction<sup>i</sup>

#### Protein-protein interaction databases

STRING<sup>i</sup> 35128.Thaps263089.

### Family & Domains<sup>i</sup>

#### Domains and Repeats

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Domain <sup>i</sup>	1 – 56	56	BRCA-2_helical  [InterPro annotation]			 Add  BLAST
Domain <sup>i</sup>	62 – 190	129	BRCA-2_OB1  [InterPro annotation]			 Add  BLAST

#### Phylogenomic databases

eggNOG<sup>i</sup> KOG4751. Eukaryota.  
ENOG410Y06W. LUCA.

InParanoid<sup>i</sup> B5YME7.

KO<sup>i</sup> K08775.

#### Family and domain databases

Gene3D <sup>i</sup>	2.40.50.140. 1 hit.
InterPro <sup>i</sup>	IPR015525. BRCA2. IPR015252. BRCA2_hlx. IPR015187. BRCA2_OB_1. IPR012340. NA-bd_OB-fold. [Graphical view]
PANTHER <sup>i</sup>	PTHR11289. PTHR11289. 1 hit.
Pfam <sup>i</sup>	PF09169. BRCA-2_helical. 1 hit. PF09103. BRCA-2_OB1. 1 hit. [Graphical view]
SUPFAM <sup>i</sup>	SSF50249. SSF50249. 1 hit. SSF81872. SSF81872. 1 hit.

# Disease databases

- OMIM (<http://www.omim.org/>)
- KEGG (<http://www.genome.jp/kegg/disease/>)
- DisGeNET  
(<http://www.disgenet.org/web/DisGeNET/menu/home>)

# OMIM

[Home](#) [About](#) [Statistics ▾](#) [Downloads ▾](#) [Help ▾](#) [External Links](#) [Terms of Use ▾](#) [Contact Us](#) [MIMmatch](#) [Donate ▾](#)



## OMIM® Online Mendelian Inheritance in Man®

An Online Catalog of Human Genes and Genetic Disorders  
Updated 29 July 2016

Search OMIM...

Search

[Advanced Search : OMIM, Clinical Synopses, Gene Map](#)

[Need help? : Example Searches, OMIM Search Help, OMIM Tutorial](#)

[Mirror site : mirror.omim.org](#)

OMIM is supported by a grant from NHGRI, licensing fees, and [generous contributions from people like you](#).



[Advanced Search](#)[Search History](#)[Display Options](#)[Retrieve Corresponding](#):[Gene Map](#)[Clinical Synopsis](#)

Search: 'alzheimer disease'

Results: 7,557 entries.

[Show 100](#) | [Download As](#)

1 : # 104300. ALZHEIMER DISEASE; AD

[ICD+](#), [Links](#)

ALZHEIMER DISEASE, FAMILIAL, 1, INCLUDED

Cytogenetic locations: 6p22.2, 7q36.1, 10q22.2, 12p13.31, 17q22, 21q21.3

Matching terms: alzheimer, disease

2 : \* 104760. AMYLOID BETA A4 PRECURSOR PROTEIN; APP

[Gene Tests](#), [Links](#)Cytogenetic location: 21q21.3, Genomic coordinates ([GRCh38](#)): 21:25,880,549-26,171,127

Matching terms: alzheimer, disease

3 : \* 104311. PRESENILIN 1; PSEN1

[Gene Tests](#), [Links](#)Cytogenetic location: 14q24.2, Genomic coordinates ([GRCh38](#)): 14:73,136,434-73,223,690

Matching terms: alzheimer, disease

4 : + 107741. APOLIPOPROTEIN E; APOE

[Gene Tests](#), [ICD+](#), [Links](#)

APOLIPOPROTEIN E, DEFICIENCY OR DEFECT OF, INCLUDED

Cytogenetic location: 19q13.32, Genomic coordinates ([GRCh38](#)): 19:44,905,748-44,909,394

Matching terms: alzheimer, disease

5 : # 607822. ALZHEIMER DISEASE 3

[ICD+](#), [Links](#)

ALZHEIMER DISEASE, FAMILIAL, 3, WITH SPASTIC PARAPARESIS AND UNUSUAL PLAQUES, INCLUDED

Cytogenetic locations: 14q24.2

alzheimer disease

[Search](#)[Advanced Search](#) | [Search History](#) | [Display Options](#)**Table of Contents for #104300**

- [Title](#)
- [Phenotype-Gene Relationships](#)
- [Text](#)
- [Description](#)
- [Clinical Features](#)
- [Other Features](#)
- [Biochemical Features](#)
- [Pathogenesis](#)
- [Inheritance](#)
- [Diagnosis](#)
- [Clinical Management](#)
- [Mapping](#)
- [Molecular Genetics](#)
- [Population Genetics](#)
- [Animal Model](#)
- [History](#)
- [Clinical Synopsis](#)
- [See Also](#)
- [References](#)
- [Contributors](#)
- [Creation Date](#)
- [Edit History](#)
- [MIMmatch \(login\)](#)

*Alternative titles: symbols***PRESENILE AND SENILE DEMENTIA**

Other entities represented in this entry:

**ALZHEIMER DISEASE, FAMILIAL, 1, INCLUDED; AD1, INCLUDED****ALZHEIMER DISEASE, EARLY-ONSET, WITH CEREBRAL AMYLOID ANGIOPATHY, INCLUDED****ALZHEIMER DISEASE, PROTECTION AGAINST, INCLUDED****Phenotype-Gene Relationships**

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
6p22.2	{Alzheimer disease, susceptibility to}	104300	AD	3	HFE	613609
7q36.1	{Alzheimer disease, late-onset, susceptibility to}	104300	AD	3	NOS3	163729
10q22.2	{Alzheimer disease, late-onset, susceptibility to}	104300	AD	3	PLAU	191840
12p13.31	{Alzheimer disease, susceptibility to}	104300	AD	3	A2M	103950
17q22	{Alzheimer disease, susceptibility to}	104300	AD	3	MPO	606989
21q21.3	Alzheimer disease 1, familial	104300	AD	3	APP	104760

[Clinical Synopsis](#)**External Links**

- [Protein](#)
- [Clinical Resources](#)
- [Animal Models](#)
- [Cell Lines](#)

- See phenotype-gene relationship

[Advanced Search ▾](#) | [Search History](#) | [Display Options ▾](#)**Table of Contents for \*613609**[Title](#)[Gene-Phenotype Relationships](#)[Text](#)[Cloning and Expression](#)[Nomenclature](#)[Biochemical Features](#)[Gene Structure](#)[Mapping](#)[Gene Function](#)[Molecular Genetics](#)[Animal Model](#)[Allelic Variants](#)[Table View](#)[See Also](#)[References](#)[Contributors](#)[Creation Date](#)[Edit History](#)[MIMmatch \(login\)](#)**\*613609****HFE GENE; HFE***Alternative titles; symbols*

HLAH

*HGNC Approved Gene Symbol: HFE**Cytogenetic location: 6p22.2      Genomic coordinates (GRCh38): 6:26,087,280-26,096,215* (from NCBI)**Gene-Phenotype Relationships**

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key
6p22.2	Hemochromatosis	235200	AR	3
	[Transferrin serum level QTL2]	614193		3
	{Alzheimer disease, susceptibility to}	104300	AD	3
	{Microvascular complications of diabetes 7}	612635		3
	{Porphyria cutanea tarda, susceptibility to}	176100	AD	3
	{Porphyria variegata, susceptibility to}	176200	AD	3

**External Links**[Genome](#)[DNA](#)[Protein](#)[Gene Info](#)[Clinical Resources](#)[Variation](#)[Animal Models](#)[Cellular Pathways](#)

Search OMIM...

Search

[Advanced Search](#) ▾

## OMIM Gene Map Statistics:

*OMIM Morbid Map Scorecard (Updated August 2nd, 2016) :*

Total number of phenotypes* for which the molecular basis is known	5,796
Total number of genes with phenotype-causing mutation	3,595

\* Phenotypes include (1) single-gene mendelian disorders and traits; (2) susceptibilities to cancer and complex disease (e.g., *BRCA1* and familial breast-ovarian cancer susceptibility, 113705.0001, and *CFH* and macular degeneration, 134370.0008); (3) variations that lead to abnormal but benign laboratory test values ("nondiseases") and blood groups (e.g., lactate dehydrogenase B deficiency, 150100.0001 and ABO blood group system, 110300.0001); and (4) select somatic cell genetic disease (e.g., *CNAS* and McCune-Albright syndrome, 139320.0008 and *IDH1* and glioblastoma multiforme, 147700.0001.)

*Dissected OMIM Morbid Map Scorecard (Updated August 2nd, 2016) :*

Class of phenotype	Phenotype	Gene*
Single gene disorders and traits	4,765	3,212
Susceptibility to complex disease or infection	700	499
"Nondiseases"	141	111
Somatic cell genetic disease	205	117

*\*Some genes may be counted more than once because mutations in a gene may cause more than one phenotype and the phenotypes may be of different classes (e.g., activating somatic *BRAF* mutation underlying cancer, 164757.0001 and germline *BRAF* mutation in Noonan syndrome, 164757.0022.)*

# KEGG database

 **KEGG DISEASE**  
Diseases viewed as perturbed states of the molecular system

\* Japanese

KEGG2 PATHWAY BRITE DISEASE DRUG ENVIRON MEDICUS Cancer Pathogen

Search DISEASE by H number, name, description, category, pathway, and  gene  
 Go

**KEGG DISEASE Database**

In KEGG, diseases are viewed as perturbed states of the molecular system, and drugs as perturbants to the molecular system. Different types of diseases, including single-gene (monogenic) diseases, multifactorial diseases, and infectious diseases, are all treated in a unified manner as follows.

```
graph TD
    GP[Genetic perturbations<br/>(germline/somatic mutations, etc.)] --> MN[Molecular network system]
    EP[Environmental perturbations<br/>(incl. pathogens, microbiome)] --> MN
    T[Therapeutic drugs] <--> MN
    GB[Genomic biomarkers] <--> MN
    DM[Diagnostic markers] --> MN
    MN --> D[Disease]
```

Our knowledge on perturbed molecular networks has been captured and represented as disease pathway maps in the KEGG PATHWAY database (see, for example, the disease pathway map of chronic myeloid leukemia hsa05220). The KEGG DISEASE database is a collection of disease entries capturing knowledge on genetic and environmental perturbations. Each disease entry is identified by the H number and contains a list of known genetic factors (disease genes), environmental factors, diagnostic markers, and therapeutic drugs (see, for example, the disease entry of chronic myeloid leukemia H00004). Diseases with known genetic factors and infectious diseases with known pathogen genomes are being organized in KEGG DISEASE.

- Human diseases [+ gene]
- Infectious diseases [+ genome]
- Human diseases in ICD-10 classification

Disease Pathway Maps and Disease Entries

[ [Brite menu](#) | [Download htext](#) ]

Human diseases

Go

▼ ▼ ▼  One-click mode

▼ **Cancers**

- ▶ Cancers of eye, brain, and central nervous system
- ▶ Cancers of the digestive system
- ▶ Cancers of haematopoietic and lymphoid tissues
- ▶ Cancers of the breast and female genital organs
- ▶ Cancers of soft tissues and bone
- ▶ Skin cancers
- ▶ Cancers of the urinary system and male genital organs
- ▶ Cancers of endocrine organs
- ▶ Head and neck cancers
- ▶ Cancers of the lung and pleura

▼ **Immune system diseases**

- ▶ Allergies and autoimmune diseases
- ▶ Primary immunodeficiency
- ▶ Other immune system diseases

▼ **Nervous system diseases**

- ▶ Neurodegenerative diseases
- ▶ Other nervous and sensory system diseases

▼ **Cardiovascular diseases**

- ▶ Cardiac diseases
- ▶ Vascular diseases
- ▶ Hematologic diseases
- ▶ Hypertensive diseases

▼ **Respiratory diseases**

- ▶ Lung diseases
- ▶ Tracheobronchial diseases

▼ **Endocrine and metabolic diseases**

- ▶ Diabetes
- ▶ Hypothalamus and pituitary gland diseases

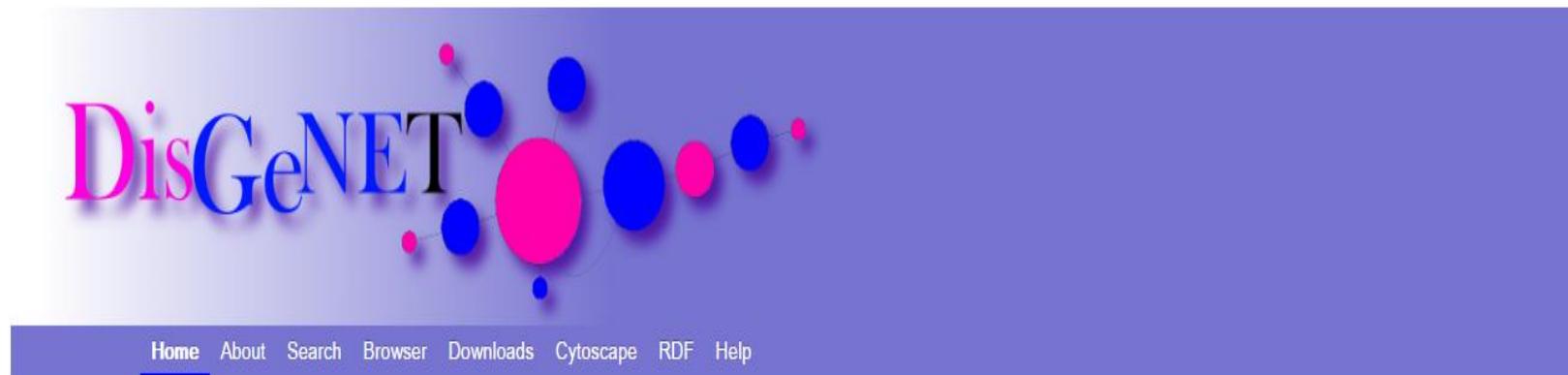
Entry	H00031	Disease
Name	Breast cancer	
Description		Breast cancer remains the most common malignancy in women worldwide and is the leading cause of cancer-related mortality. More than 1-2 million cases are diagnosed every year, affecting 10-12% of the female population and accounting for 500 000 deaths per year worldwide. Approximately 5-10% are thought to be inherited. The hereditary breast cancer syndrome includes genetic alterations in various susceptibility genes such as p53, PTEN, BRCA1, and BRCA2. Sporadic breast cancers result from a serial stepwise accumulation of acquired and uncorrected mutations in somatic genes, without any germline mutation playing a role. Oncogenes that have been reported to play an early role in sporadic breast cancer are MYC, CCND1 (Cyclin D1) and ERBB2 (HER2/neu). In sporadic breast cancer, mutational inactivation of BRCA1/2 is rare. However, non-mutational functional suppression could result from various mechanisms, such as hypermethylation of the BRCA1 promoter.
Category	Cancer	
Brite		<p>Human diseases [BR:br08402]</p> <p>Cancers</p> <p>Cancers of the breast and female genital organs</p> <p>H00031 Breast cancer</p> <p>Human diseases in ICD-10 classification [BR:br08403]</p> <p>2. Neoplasms (C00-D48)</p> <p>C50-C50 Malignant neoplasm of breast</p> <p>C50 Malignant neoplasm of breast</p> <p>H00031 Breast cancer</p> <p>Tumor markers [br08442.html]</p> <p>Commonly used tumor markers</p> <p>H00031</p> <p>Cancer-associated carbohydrates [br08441.html]</p> <p>H00031</p> <p><a href="#">BRITE hierarchy</a></p>
Pathway	hsa05206	MicroRNAs in cancer
Gene		<p>BRCA1 (germline mutation, hypermethylation) [HSA:672] [KO:K10605]</p> <p>BRCA2 (germline mutation) [HSA:675] [KO:K08775]</p> <p>p53 (mutation) [HSA:7157] [KO:K04451]</p>

## All links

- Ontology (2)
  - KEGG BRITE (2)
  - Pathway (2)
    - KEGG PATHWAY (2)
- Drug (10)
  - KEGG DRUG (4)
  - KEGG DGROUP (6)
- Chemical substance (1)
  - KEGG GLYCAN (1)
- Gene (25)
  - KEGG ORTHOLOGY (9)
  - KEGG GENES (16)
- Literature (10)
  - PubMed (10)
- All databases (50)

[Download RDF](#)

# DisGeNET database



One of the most challenging problems in biomedical research is to understand the underlying mechanisms of complex diseases. Great effort has been spent on finding the genes associated to diseases (Botstein and Risch, 2003; Kann, 2009). However, more and more evidences indicate that most human diseases cannot be attributed to a single gene but arise due to complex interactions among multiple genetic variants and environmental risk factors (Hirschhorn and Daly, 2005). Several databases have been developed storing associations between genes and diseases such as CTD™ (Davis, *et al.*, 2014), OMIM® (Hamosh *et al.*, 2005) and the NHGRI-EBI GWAS catalog (Welter *et al.*, 2014). Each of these databases focuses on different aspects of the phenotype-genotype relationship, and due to the nature of the database curation process, they are not complete. Hence, integration of different databases with information extracted from the literature is needed to allow a comprehensive view of the state of the art knowledge within this research field. With this need in mind, we have created DisGeNET.

DisGeNET is a discovery platform integrating information on gene-disease associations (GDAs) from several public data sources and the literature (Piñero *et al.*, 2015). The current version contains (DisGeNET v4.0) contains 429,036 associations, between 17,381 genes and 15,093 diseases, disorders and clinical or abnormal human phenotypes. Given the large number of GDAs compiled in DisGeNET, we have also developed a score in order to rank the associations based on the supporting evidence. Importantly, useful tools have also been created to explore and analyze the data contained in DisGeNET. DisGeNET can be queried through Search and Browse functionalities available from this web interface, or by a plugin created for Cytoscape to query and analyze a network representation of the data. Moreover, DisGeNET data can be queried by downloading the SQLite database to your local repository. Furthermore, an RDF (Resource Description Framework) representation of DisGeNET database is also available. It can be queried using our SPARQL endpoint and a Faceted Browser. Follow the link for more information.

DisGeNET database has been cited by several papers. Some of them can be reviewed here.

The DisGeNET database is made available under the Open Database License. Any rights in individual contents of the database are licensed under the Database Contents License.

## Tweets by @DisGeNET

DisGeNET  
@DisGeNET

The registration deadline for  
DisGeNET tutorial at  
#eccb2016 is August 19  
eccb2016.org/programme/tuto  
...

23h

DisGeNET Retweeted

eccb2016  
@eccb2016

If you did not do so yet: register  
now for the #eccb2016! Online  
registration deadline: August  
19. eccb2016.org  
#bioinformatics

# DisGeNET



Home   About   Search   Browser   **Downloads**   Cytoscape   RDF   Help

#### DATA IN TAB SEPARATED FILES (VERSION 4.0)

1. File with CURATED gene-disease associations
2. File with BeFree gene-disease associations
3. File with ALL gene-disease associations in DisGeNET
4. File with SNP-gene-disease associations found by BeFree System
5. File with SNP-gene-disease associations in DisGeNET
6. Publications with gene-disease associations from BeFree System
7. README

#### RDF LINKED DATASET (VERSION 4.0)

1. RDF Downloads

#### NANOPUBLICATIONS LINKED DATASET (VERSION 4.0)

1. DisGeNET Nanopublications dataset v4.0.0.0

#### MAPPINGS

1. UniProt Downloads
2. UMLS CUI to MeSH Identifier

#### SCRIPTS

# Protein databases

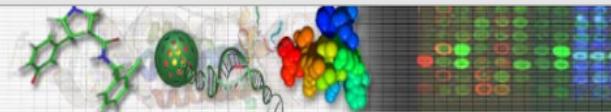
- HPRD (<http://www.hprd.org/>)
- DIP (<http://dip.doe-mbi.ucla.edu/dip/Main.cgi>)
- PDB (<http://www.wwpdb.org/>)

# HPRD

← → C hprd.org/index\_html



You are at: HPRD



Query

Browse

Blast

FAQs

Download

Human Proteinpedia

Pathways

PhosphoMotif Finder

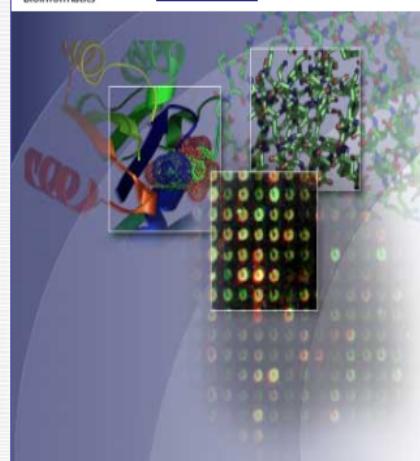
Become a "Molecule Authority"

## News

["Human Proteinpedia enables data sharing of human proteins" in February 2008 issue of \*Nature Biotechnology\*](#)

[PhosphoMotif Finder, published in February 2007 issue of \*Nature Biotechnology\*](#)

[Comparison of Protein-Protein Interaction Databases, published in \*BMC Bioinformatics\*](#)



## Highlights

### PhosphoMotif Finder

Allows you to check if your protein contains any phosphorylation motif described in the literature

### Pathways

A set of 25 curated signaling pathways are available as part of a new pathway resource that we have developed called 'NetPath.'

### HPRD Release 9 New

The latest Release 9 is available for download. [Click here ...](#)

### Search by PubMed New

## Statistics

Protein Entries	30,047
Protein-Protein Interactions	41,327
PTMs	93,710
Protein Expression	112,158
Subcellular Localization	22,490
Domains	470
PubMed Links	453,521

## About HPRD

COMMERCIAL ENTITIES MAY NOT USE THIS SITE WITHOUT PRIOR LICENSING AUTHORIZATION. PLEASE SEND AN E-MAIL FOR FURTHER INFORMATION ABOUT LICENSING.

The Human Protein Reference Database represents a centralized platform to visually depict and integrate information pertaining to domain architecture, post-translational modifications, interaction networks and disease association for each protein in the human proteome. All the information in HPRD has been manually extracted from the literature by expert biologists who read, interpret and analyze the published data. HPRD has been created using an object oriented database in Zope, an open source web application server, that provides versatility in query functions and allows data to be displayed dynamically.

# Query

 Human Protein Reference Database

You are at: HPRD >> Query

**Query**

The default behavior if more than one term is entered within a field is 'AND.' e.g. entering 'SH2 SH3' in 'Domain' search field will search for all the proteins that have both SH2 and SH3 domains. Similarly, if more than one field is filled in, it will be treated as an 'AND' query. For more information go to the [FAQ](#).

Protein Name

Accession Number RefSeq

HPRD Identifier

Gene Symbol

Chromosome Locus

Molecular Class  [See List](#)

PTMs  [See List](#)

Cellular Component  [See List](#)

Domain Name  [See List](#)

Motif  [See List](#)

Expression  [See List](#)

Length of Protein Sequence From:  to:  in amino acids

Molecular Weight From:  to:  in kDa

Diseases  breast cancer

Please send any questions or comments about the Human Protein Reference Database to [help](#).

Copyright © Johns Hopkins University and the Institute of Bioinformatics.

This is a joint project between:

 PandeyLab and  Institute of Bioinformatics



You are at: HPRD >> Query >> Query Result

Query

Browse

Blast

FAQs

Download

Human Proteinpedia

Pathways

PhosphoMotif Finder

Become a "Molecule Authority"

## Query Results

Your search for **breast cancer** in Disease reports **17** matches.

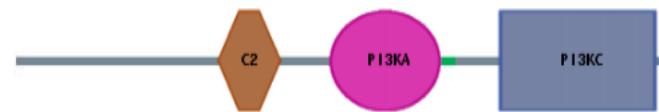
Results are sorted by relevance, indicating the number of occurrences of items searched for.

Showing matches **1** to **17**

Sort by Relevance | Number of PTMs | Protein Length | Number of Interactions

- 1 Name : [Phosphatidylinositol 3 kinase, catalytic subunit alpha](#)

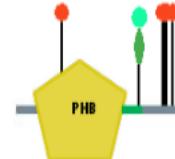
Molecule Function : Lipid kinase activity



Number of Interactions : 20

- 2 Name : [Prohibitin](#)

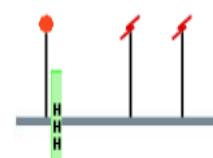
Molecule Function : Receptor signaling complex scaffold activity



Number of Interactions : 15

- 3 Name : [RAD51](#)

Molecule Function : DNA binding



[Query](#)

[Browse](#)

[Blast](#)

[FAQs](#)

[Download](#)

[Human Proteinpedia](#)

[Pathways](#)

[PhosphoMotif Finder](#)

[Become a Molecule Authority](#)

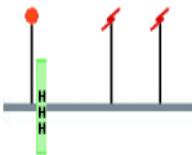
# RAD51

Molecular Class: DNA binding protein

Molecular Function: DNA binding

Biological Process: Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism

## Isoform 1



GO TO: Isoform\_1 ▾

[ALTERNATE NAMES](#)   [DISEASES](#)   [PTMs & SUBSTRATES](#)

SUMMARY   [SEQUENCE](#)   [INTERACTIONS](#)   [EXTERNAL LINKS](#)

### General

HPRD ID: 01557  
Gene Symbol: [RAD51](#)

Molecular Weight (Da): 36780  
Gene Map Locus: 15q15.1

### Localization

Primary

[Nucleus](#)

Alternate

### Domains and Motifs

Domains  
HHH 58 - 77

Motifs

### Expression

Site of Expression  
Leukocyte  
Testis

# Browse

← → ⌂ hprd.org/browse/localizations

Human Protein Reference Database

You are at: HPRD >> Browse >> Localization

Query    Browse    Blast    FAQs    Download    Human Proteinpedia    Pathways    PhosphoMotif Finder    Become a 'Molecule Authority'

## Browse

MOLECULE CLASS    DOMAINS    MOTIFS    PTMs    LOCALIZATIONS

Please select localization class to browse :

<a href="#">Acrosome</a>	<a href="#">Golgi membrane</a>	<a href="#">Synapse</a>
<a href="#">Actin cytoskeleton</a>	<a href="#">Golgi vesicle</a>	<a href="#">Synaptic vesicle</a>
<a href="#">Actin filament</a>	<a href="#">Immunological synapse</a>	<a href="#">Tubulin</a>
<a href="#">Apical membrane</a>	<a href="#">Integral to membrane</a>	<a href="#">Unknown membrane</a>
<a href="#">Axon</a>	<a href="#">Integral to plasma membrane</a>	<a href="#">Vesicle</a>
<a href="#">Axoneme</a>	<a href="#">Intermediate filament</a>	<a href="#">Vesicular fraction</a>
<a href="#">Azurophil granule</a>	<a href="#">Intracellular vesicle</a>	<a href="#">Zymogen granule</a>
<a href="#">Basolateral membrane</a>	<a href="#">Kinetochore</a>	
<a href="#">Brush border</a>	<a href="#">Late endosome</a>	
<a href="#">Caveola</a>	<a href="#">Lysosome</a>	
<a href="#">Cell junction</a>	<a href="#">Membrane fraction</a>	
<a href="#">Cell projection</a>	<a href="#">Microsome</a>	
<a href="#">Cell surface</a>	<a href="#">Microtubule</a>	
<a href="#">Centriole</a>	<a href="#">Microtubule cytoskeleton</a>	
<a href="#">Centrosome</a>	<a href="#">Mitochondrial intermembrane space</a>	
<a href="#">Chromaffin granule</a>	<a href="#">Mitochondrial matrix</a>	
<a href="#">Chromosome</a>	<a href="#">Mitochondrial membrane</a>	
<a href="#">Cilium</a>	<a href="#">Mitochondrion</a>	
<a href="#">Clathrin-coated vesicle</a>	<a href="#">Myosin</a>	
<a href="#">Coated vesicle</a>	<a href="#">Nuclear matrix</a>	
<a href="#">Cytoplasm</a>	<a href="#">Nuclear membrane</a>	
<a href="#">Cytoplasmic vesicle</a>	<a href="#">Nucleolus</a>	
<a href="#">Cytoplasmic vesicle membrane</a>	<a href="#">Nucleoplasm</a>	
<a href="#">Cytoskeleton</a>	<a href="#">Nucleus</a>	
<a href="#">Cytosol</a>	<a href="#">Perinuclear region</a>	
<a href="#">Dendrite</a>	<a href="#">Perinuclear vesicle</a>	
<a href="#">Dense-core vesicle</a>	<a href="#">Peroxisomal matrix</a>	
<a href="#">Desmosome</a>	<a href="#">Peroxisomal membrane</a>	
<a href="#">Early endosome</a>	<a href="#">Peroxisome</a>	
<a href="#">Endocytic vesicle</a>	<a href="#">Plasma membrane</a>	
<a href="#">Endoplasmic reticulum</a>	<a href="#">Postsynaptic density</a>	
<a href="#">Endoplasmic reticulum lumen</a>	<a href="#">Ribosome</a>	
<a href="#">Endoplasmic reticulum membrane</a>	<a href="#">Rough microsome</a>	
<a href="#">Endosome</a>	<a href="#">Sarcoplasm</a>	
<a href="#">Extracellular</a>	<a href="#">Sarcoplasmic reticulum</a>	
<a href="#">Extracellular matrix</a>	<a href="#">Secreted</a>	
<a href="#">Extracellular space</a>	<a href="#">Secretory granule</a>	
<a href="#">Focal adhesion</a>	<a href="#">Secretory vesicle</a>	
<a href="#">Golgi apparatus</a>	<a href="#">Smooth microsome</a>	
<a href="#">Golgi lumen</a>	<a href="#">Soluble fraction</a>	



You are at: HPRD > browse

Query

Browse

Blast

FAQs

Download

Human  
Proteinpedia

Pathways

PhosphoMotif  
Finder

Become a  
"Molecule Authority"

## Browse Results

Your search for Protein(s) Cellular Component at the "Nucleolus", reports 697 matches.  
Results are sorted by relevance, indicating the number of occurrences of items searched for.  
Showing matches 1 to 30

↓ Sort by

Relevance

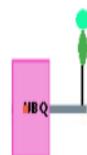
Number of PTMs

Protein Length

Number of Interactions

1 Name : FAU

Molecule Class : Ubiquitin proteasome system protein



Number of Interactions : 2

2 Name : Replication factor C4

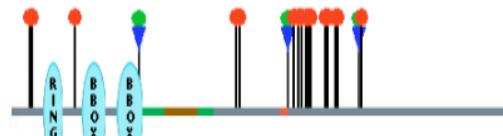
Molecule Class : DNA binding protein



Number of Interactions : 18

3 Name : PML

Molecule Class : Transcription regulatory protein



# Download

 Human Protein Reference Database

You are at HPRD

**Download page**

**Comparison of Protein-Protein Interaction databases, published in BMC Bioinformatics**

Please fill in the following details to download

**FILES AVAILABLE FOR DOWNLOAD**

File name	File description	File type	Version number	Date created
HPRD_Release9_041310.tar.gz	This file contains human binary protein-protein interactions in tab delimited format.	Tab delimited	Release 9	04-13-10 Apr 13, 2010
HPRD_FLAT_FILES_041310.tar.gz	This file contains all entries contained in HPRD including features of proteins such as post-translational modifications, tissue expression, subcellular localization and protein-protein interactions in tab delimited file format as per the users request.	Tab delimited	Release 9	04-13-10 Apr 13, 2010
HPRD_XML_041310.tar.gz	This file contains all entries contained in HPRD including features of proteins such as post-translational modifications, tissue expression, subcellular localization and protein-protein interactions etc.	XML	Release 9	04-13-10 Apr 13, 2010
HPRD_PSIMI_041310.tar.gz	This file contains only protein-protein interactions of all entries contained in HPRD. The data is in PSI-MI format version 2.5	XML	Release 9	04-13-10 Apr 13, 2010
HPRD_SINGLE_PSIMI_041310.xml.tar.gz	This file contains only protein-protein interactions but all interaction records are concatenated into a single XML file. The data is in PSI-MI format version 2.5	XML	Release 9	04-13-10 Apr 13, 2010

**ACADEMIC USERS**

Name

Email

Designation

Institution

Address

Which features of HPRD will be most useful to you

Are you the Principal Investigator?  No  Yes

If you are not a Principal Investigator, please provide details of the Principal Investigator or Laboratory Chief.

Name

Institution



# DIP



User: | Sign up | Log in

ProLinks | Help | Contact Us

DIP Data DIP Tools About DIP News Resources Preferences



Contribute to DIP!  
[Submit your interactions](#)



View interaction networks  
[Get software](#)



Get all interactions in DIP  
[Download DIP data](#)

## SEARCH DIP

Search DIP

[Help: Tutorial](#)

Enter names, keywords, or identifiers associated with a protein, gene, interaction, or publication. (See all [identifier types](#) recognized by DIP.)

For more search options, try [DIP Search](#)

## DIP News

Recent DIP news:

2015-01-01  
Yet another interaction dataset...

2014-02-04  
Updated MiSink plugin released

2014-01-17  
More interaction data

[See more DIP News](#)

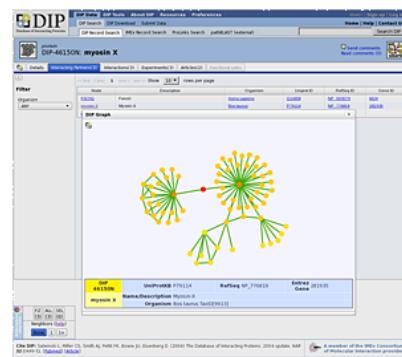
## Search Help

Example DIP searches:

- Protein name: [MVP](#)
- Gene name: [SUP35](#)
- Keywords: [yeast actin](#)
- UniProt ID: [P52960](#)
- PubMed ID: [8676499](#)
- PDB ID: [1PQS](#)

[More help with DIP Searches](#)

## Welcome



Welcome to the new and, hopefully, improved DIP Database site. While retaining most of the functionalities of the legacy site, the new portal utilizes modern web technologies to offer new graphical user interface while continuing tight integration with Cytoscape. Additional features include direct access to MIQL/PSICQUIC search engine site customization for registered users. Next few months will bring reimplementations of the remaining functionalities offered by the legacy site and fine-tuning of the caching subsystem which will make the site more responsive when dealing with large interaction datasets.

NOTE: If you prefer to visit the old site, please, follow to:

<http://dip.doe-mbi.ucla.edu>

## DIP Contents

DIP contains information on:

- 28,764 proteins
- 81,827 interactions
- 8,187 articles
- 826 organisms

It is based on:

- 81,635 experiments
- 29,537 automated inferences [2]
- 6,273 authors' inferences [2]

[See more DIP Statistics](#)

# PDB

WORLDWIDE  
**wwPDB**  
PROTEIN DATA BANK

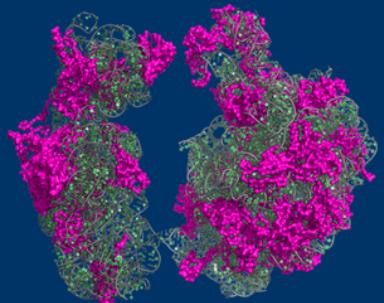
VALIDATION ▾ DEPOSITION ▾ DATA DICTIONARIES ▾ DOCUMENTATION ▾ TASK FORCES ▾ STATISTICS ▾ ABOUT ▾

wwPDB Foundation

Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

The Worldwide PDB (wwPDB) organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.

Learn more about PDB **HISTORY** and **FUTURE**.



**Validate Structure**  
or View validation reports

**Deposit Structure**  
All Deposition Resources

**Download Archive**  
Instructions

### wwPDB Members

wwPDB data centers serve as deposition, annotation, and distribution sites of the PDB archive. Each site offers tools for searching, visualizing, and analyzing PDB data.

**PDBe**  
Protein Data Bank in Europe 

Rich information about all PDB entries, multiple search and browse facilities, advanced services including PDBePISA, PDBeFold and PDBeMotif, advanced visualisation and validation of NMR and EM structures, tools for bioinformaticians.

**RCSB PDB**  
Research Collaboratory for Structural Bioinformatics Protein Data Bank 

Simple and advanced searching for macromolecules and ligands, tabular reports, specialized visualization tools, sequence-structure comparisons, RCSB PDB Mobile, Molecule of the Month and other educational resources at PDB-101, and more.

**PDBj**

### wwPDB Resources

**Data Dictionaries**

- Macromolecular Dictionary (PDBx/mmCIF)
- Small Molecule Dictionary (CCD)
- Peptide-like antibiotic and inhibitor molecules (BIRD)

**Annotation**

- Procedures and policies
- Improvements for consistency and accuracy

**Community Input:**  
Task Forces and Working Groups

- Validation Task Forces (X-ray, NMR, 3DEM)
- Small Angle Scattering Task Force
- PDB/mmCIF Working Group
- Hybrid/Integrative Methods Task Force
- Ligand Validation Workshop

**PDB Data Growth & Usage Statistics**

- Depositions: by data center, by year, and by depositor location
- Downloads: by year for all entries

### News & Announcements

**07/06/2016**

Announcement: Map Volume Deposition to EMDB Will Be Mandatory for PDB Depositions of 3DEM models Starting September 6th, 2016

Effective September 6th, 2016, deposition to the PDB of atomic models determined by 3D Electron cryo-Microscopy (3DEM) will require prior or simultaneous deposition of the associated 3DEM volume maps to EMDB.

[Read more](#)

**06/23/2016**

Deposit ORCID and Grant Information with PDB Data

To enable better annotation and tracking, the wwPDB partners encourage depositors to provide ORCID identifiers and information on relevant grants funding their research when depositing PDB data.

[Read more](#)

28

# lncRNA databases

- LncRNADisease database  
(<http://www.cuilab.cn/lncrnadisease>)
- Lncrnadb (<http://lncrnadb.org/>)
- LNCipedia.org (<http://www.lncipedia.org/>)

# LncRNADisease database

## The LncRNA and Disease Database

[Home](#)   [Browse](#)   [Search](#)   [Interaction](#)   [Predict](#)   [Download](#)  
[Submit](#)   [Help](#)

### Welcome to the LncRNADisease database

Long noncoding RNAs (lncRNAs) represent the RNAs that have a size of >200 nucleotides and lack the capability of protein-coding. Although lncRNAs are among one of the least well-understood molecules, emerging studies have revealed that lncRNAs play critical roles in a broad range of biological processes and are associated with a number of diseases, i.e. cancer, cardiovascular disease, and neurodegeneration disease. lncRNAs are becoming critically important for the understanding of life sciences, especially diseases. The [LncRNADisease](#) database is not only a resource that curated the experimentally supported lncRNA-disease association data but also a platform that integrated tool(s) for predicting novel lncRNA-disease associations. In addition, LncRNADisease also curated lncRNA interactions in various levels, including protein, RNA, miRNA, and DNA.

Currently, users can

- browse the experimentally supported lncRNA-disease association data;
- search the experimentally supported lncRNA-disease association data;
- browse the experimentally supported lncRNA interaction data;
- predict potential associated diseases for a novel lncRNA based on its genomic context;
- download the experimentally supported lncRNA-disease association data;
- download the experimentally supported lncRNA interaction data;

**Statistics:**  
LncRNADisease database integrated more than 1000 lncRNA-disease entries and 475 lncRNA interaction entries, including 321 lncRNAs and 221 diseases from ~500 publications. LncRNADisease also provided the predicted associated diseases of 1564 human lncRNAs.

**Links:**  
[IncRNAb](#): a reference database for long noncoding RNAs.  
[NRED](#): a database of long noncoding RNA expression.  
[NONCODE](#): an integrated knowledge database of non-coding RNAs.  
[HMDD](#): a human microRNA disease database.  
[OMIM](#): a database of human genes and genetic disorders.  
[GAD](#): a Genetic Association Database.

**History:**  
Jun-2, 2015, the LncRNADisease database was updated with 74 lncRNA-disease entries.  
Jun-11, 2014, the LncRNADisease database was updated with 267 lncRNA-disease entries.  
Jan-18, 2014, the LncRNADisease database was updated with 110 lncRNA-disease entries.  
Dec-3, 2013, the LncRNADisease database was updated with 46 lncRNA-disease entries.

# The LncRNA and Disease Database

Home

Browse

Search

Interaction

Predict

Download

Submit

Help

You can search the entries by such keywords:

LncRNA ▾  
LncRNA  
Disease

[Click to Search](#)

*Begin a new search!*

# The LncRNA and Disease Database

Home    Browse    **Search**    Interaction    Predict    Download

Submit    Help

You can search the entries by such keywords:

Disease ▾        [Click to Search](#)

[Reset all](#)

LncRNA name	Disease name	Dysfunction type	Description	Chr	Start	End	Strand	Species	Alias	Genbank	Sequence	Reference
BCAR4	breast cancer	Expression	BCAR4 is expressed in 27% of primary breast tumors. Forced expression of BCAR4 in human ZR-75-1 and MCF7 breast cancer cells resulted in cell proliferation in the absence of estrogen and in the presence of various antiestrogens. BCAR4 may be a good target for treating antiestrogen-resistant breast cancer.	chr16	11913687	11922689	-	Human	BCAR4	<a href="#">NR_024049.1</a>	<a href="#">Gene / RNA</a>	<a href="#">21506106</a>
BCAR4	breast cancer	Expression	High BCAR4 mRNA levels were associated with poor MFS and overall survival, reflecting tumour aggressiveness.	chr16	11913687	11922689	-	Human	BCAR4	<a href="#">NR_024049.1</a>	<a href="#">Gene / RNA</a>	<a href="#">20859285</a>
			Breast cancer anti-estrogen resistance 4									

# The LncRNA and Disease Database

[Home](#)[Browse](#)[Search](#)[Interaction](#)[Predict](#)[Download](#)[Submit](#)[Help](#)

The experimentally supported lncRNA-disease association data ([txt](#) and [xls](#)).

The experimentally supported lncRNA interaction data ([txt](#) and [xls](#)).

The predicted lncRNA-disease association data ([txt](#)).

The gene sequence of lncRNA ([txt](#)).

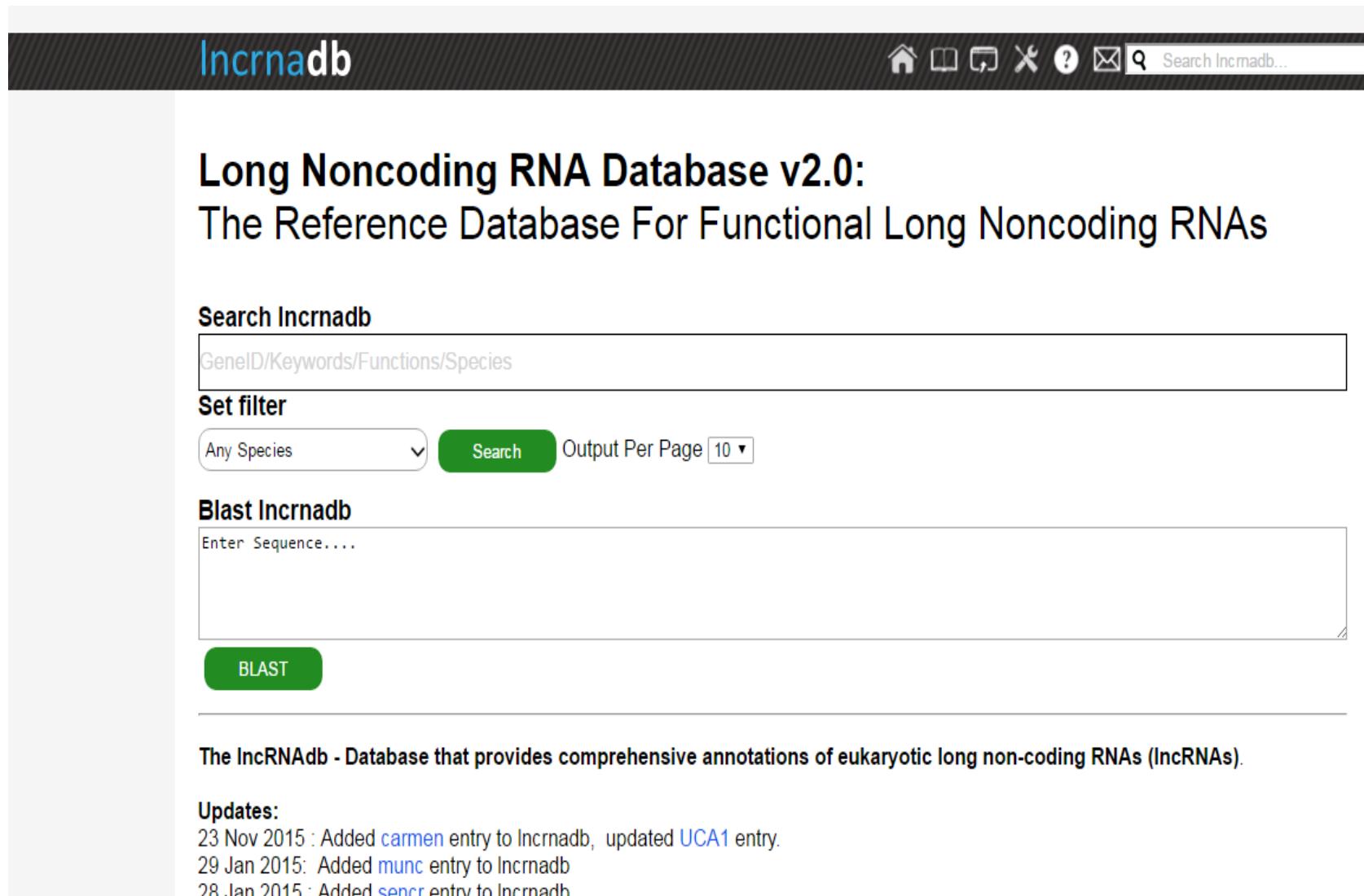
The RNA sequence of lncRNA ([txt](#)).

A list of lncRNAs predicted to be associated with cardiovascular function and disease ([rar](#)).

Tissue specific lincRNAs ([txt](#)).

LincRNAs associated disease predicted by co-expression analysis ([xls](#)).

# Incrnadb database



The screenshot shows the Incrnadb database homepage. At the top, there is a navigation bar with icons for home, search, and help. Below the header, the title "Long Noncoding RNA Database v2.0: The Reference Database For Functional Long Noncoding RNAs" is displayed. The main content area contains two search sections: "Search Incrnadb" and "Blast Incrnadb". The "Search Incrnadb" section includes a search input field for "GenelD/Keywords/Functions/Species", a dropdown for "Any Species", a "Search" button, and a "Output Per Page" dropdown set to 10. The "Blast Incrnadb" section has a text input field for "Enter Sequence...." and a "BLAST" button. A descriptive text at the bottom states: "The IncRNAdb - Database that provides comprehensive annotations of eukaryotic long non-coding RNAs (lncRNAs)." A "Updates:" section lists recent additions: "23 Nov 2015 : Added [carmen](#) entry to Incrnadb, updated [UCA1](#) entry.", "29 Jan 2015: Added [munc](#) entry to Incrnadb", and "28 Jan 2015 : Added [senor](#) entry to Incrnadb".

**Incrnadb**

Long Noncoding RNA Database v2.0:  
The Reference Database For Functional Long Noncoding RNAs

Search Incrnadb

GenelD/Keywords/Functions/Species

Set filter

Any Species  Output Per Page 10

Blast Incrnadb

Enter Sequence....

The IncRNAdb - Database that provides comprehensive annotations of eukaryotic long non-coding RNAs (lncRNAs).

**Updates:**

23 Nov 2015 : Added [carmen](#) entry to Incrnadb, updated [UCA1](#) entry.

29 Jan 2015: Added [munc](#) entry to Incrnadb

28 Jan 2015 : Added [senor](#) entry to Incrnadb

# Results

## Full text search

Homo sapiens

## Set filter

Any Species

Search

Output Per Page 10

## Blast Incrnadb

### Results

Displaying results 1 to 10 of 184

Showing 10

[Download Results as XML](#)

#### 1.Air - Alias : Airn, IGF2RAS, ENSG00000268257

Description : Antisense Igf2r RNA

Species : Homo sapiens, Mus musculus

#### 2.BC200 - Alias : BCYRN1, brain cytoplasmic RNA 1

Description : 200 nucleotide ncRNA ((Tiedge 1993)) exapted from an Alu element ((Watson 1987)). Transcribed by RNA polymerase III ((Martignetti 1993)). Three structural domains, 5' that shares homology with Alu elements, a central A rich region and a 3' unique region ((Tiedge 1993)).

Species : Papio hamadryas, Macaca fascicularis, Gorilla gorilla, Saguinus imperator, Hylobates lar, Pongo pygmaeus, Pan troglodytes, Saguinus oedipus, Macaca mulatta, Cercopithecus aethiops (Green monkey), Pan paniscus (Pygmy chimpanzee), Homo sapiens, Aotus trivirgatus

#### 3.ANRIL - Alias : CDKN2B-AS1, p15AS, ENSG00000240498

Description : Present in multiple splicing isoforms ((Folkerson 2009)) (such as the CDKN2B RefSeq sequence NM\_003520.3 of ~3.9kb). Also

[login | register](#)

# LNCipedia.org v. 4.0

A comprehensive compendium of human long non-coding RNAs

**HOME** **SEARCH** **GENOME** **SUBMIT** **DOWNLOAD** **ABOUT** **CONTACT**

## WELCOME

May 25, 2016

Welcome to LNCipedia 4.0, the latest version of this long non-coding RNA database contains 118,777 human annotated lncRNAs.

### Introduction

Long non-coding RNA (lncRNA) constitutes a large and diverse class of non-coding RNA genes. While several lncRNAs have been functionally annotated, the majority remains to be characterized. Different high-throughput methods to identify new lncRNAs (including RNA sequencing and annotation of chromatin-state maps) have been applied in various studies resulting in multiple unrelated lncRNA datasets.

Lncipedia.org is an integrated database of 118,777 human annotated lncRNA transcripts obtained from different sources. In addition to basic transcript information and structure, several statistics are calculated for each entry in the database, such as secondary structure information, protein coding potential and microRNA binding sites.

The database is publicly available and allows users to query and download lncRNA sequences and structures based on different search criteria. The database may serve as a source of information on individual lncRNAs or as a starting point for large-scale studies.

[HOME](#)[SEARCH](#)[GENOME](#)[SUBMIT](#)[DOWNLOAD](#)[ABOUT](#)[CONTACT](#)

## SEARCH

**Name/ID:****Source:** all**Chromosomal location (hg19):** - :  - **Class:****Keyword(s):****(Partial) sequence:** Coding potential ID history converter[SEARCH](#)[EXPORT ▾](#)

118485 lncRNA transcripts found (5925 pages)

Transcript ID	Gene ID	Location (hg19)	strand	transcript size
lnc-FAM72C-2:1	lnc-FAM72C-2	chr1:149590752-149591762	-	398
FAM138A:3	FAM138A	chr1:35245-36073	-	590