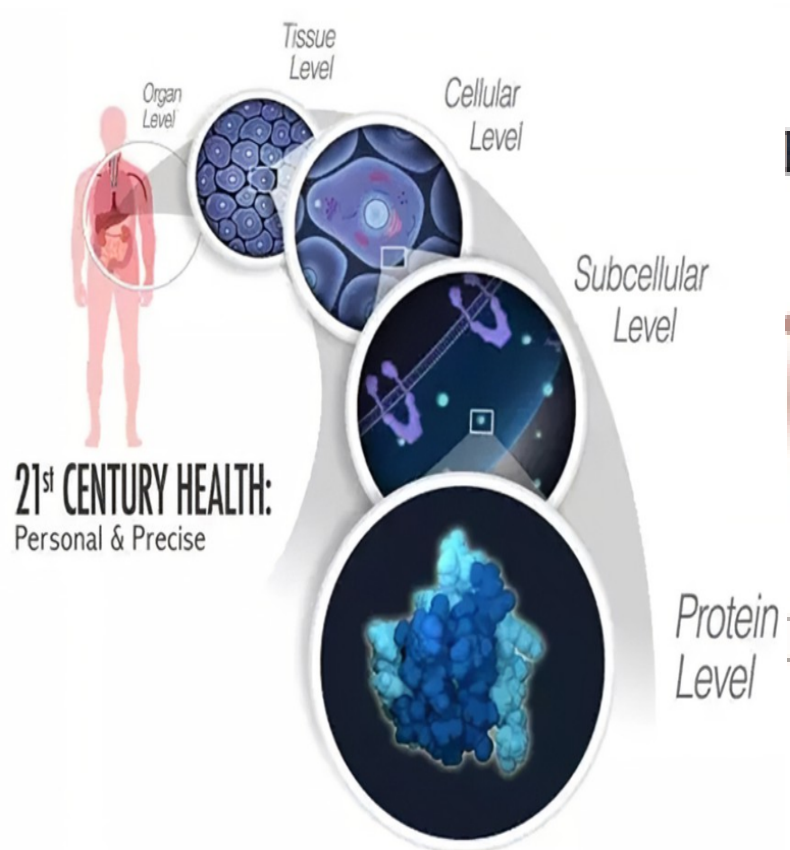
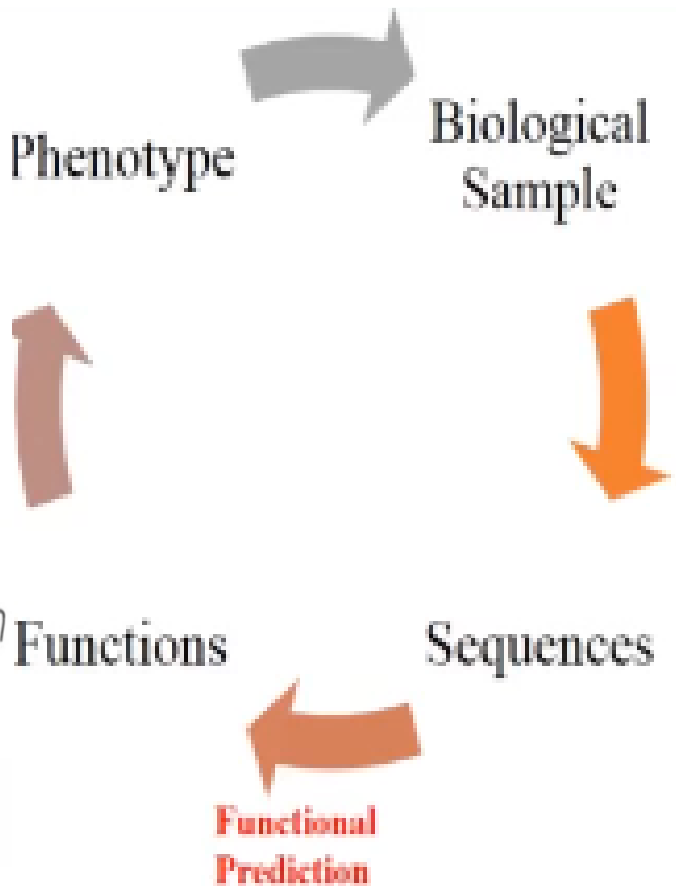
基因蛋白功能查询

张程祎 22.6.16

蛋白组学





- Proteins are the molecules that conduct the cell's business. The genes of every cell in your body are identical, but the proteome of a neuron doesn't resemble that of an adipocyte.
- The proteome provides 1000-fold more cellular information than the genome: there are ~23,000 human genes, ~100,000 transcripts, and over 20 million protein variants
- Because of transcriptional and translational levels of cellular control, not every gene is transcribed, and not every RNA is translated. Conversely, stable proteins often outlive transcripts from which they were made.
- Proteins are preferred targets for therapeutic agents and diagnostic tests. Immunoassays are quick, easy, capable for point-of-care use, and cost effective

金开瑞讲座

蛋白组学

蛋白质组学(Proteomics)是研究细胞、组织或生物体中蛋白质组成、定位、变化及其相互作用规律的科学，包括对蛋白质表达模式和蛋白质组功能模式的研究。蛋白质组学的发展对寻找疾病的诊断标志、筛选药物靶点、毒理学研究等有重要意义，也因此被广泛应用于医学研究。

➤目标

①细胞中蛋白质的含量

②定位

③活性

④修饰

➤方法

①蛋白质双向电泳

②氨基酸序列测定（包括N端测序和C端测序）

③质谱

④**生物信息学**

未知蛋白功能注释

功能预测

- **基础知识**
  - 基本假设：序列一级结构相似 → 功能相似
  - biomart R语言软件包，基因功能查询
- **通用数据库**
  - GO（分子功能、细胞定位、生物过程）
  - KEGG（代谢途径）
  - COG

https://www.jianshu.com/p/b4516bc31bfd
https://www.jianshu.com/p/48716fa7321b
https://www.jianshu.com/p/b38bbeea4223

已知蛋白功能查询

- 差异蛋白功能富集
- 关键蛋白功能查询

Functional Enrichment

Enrichment Analysis

Why we need enrichment?

Many functional nodes would be gathered and overlap if just annotate genes/proteins directly, which may puzzle researchers. So we hope to filter and screen it to achieve more significative functional nodes.

How to achieve enrichment?

- Fisher's exact test
- Cumulative supper hypergeometric test

Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.

Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles

Aravind Subramanian[a,b], Pablo Tamayo[a,b], Vamsi K. Mootha[a,c], Sayan Mukherjee[d], Benjamin L. Ebert[a,e], Michael A. Gillette[a,f], Amanda Paulovich[g], Scott L. Pomeroy[h], Todd R. Golub[a,e], Eric S. Lander[a,c,i,j,k], and Jill P. Mesirov[a,b]

[a]Broad Institute of Massachusetts Institute of Technology and Harvard, 320 Charles Street, Cambridge, MA 02141; [b]Department of Systems Biology, Albert 536, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02446; [c]Institute for Genome Sciences and Policy, Center for Interdisciplinary Engineering, Medicine, and Applied Sciences, Duke University, 101 Science Drive, Durham, NC 27708; [e]Department of Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; [f]Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114; [g]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, C2-023, P.O. Box 19024, Seattle, WA 98109-1024; [h]Department of Neurology, Enders 260, Children's Hospital, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115; [i]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142; and [j]Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA 02142

Contributed by Eric S. Lander, August 2, 2005

Although genomewide RNA expression analysis has become a routine tool in biomedical research, extracting biological insight from such information remains a major challenge. Here, we describe a powerful analytical method called Gene Set Enrichment Analysis (GSEA) for interpreting gene expression data. The method derives its power by focusing on gene sets, that is, groups of genes that share common biological function, chromosomal location, or regulation. We demonstrate how GSEA yields insights into several cancer-related data sets, including leukemia and lung cancer. Notably, where single-gene analysis finds little similarity between two independent studies of patient survival in lung cancer, GSEA reveals many biological pathways in common. The GSEA method is embodied in a freely available software package, together with an initial database of 1,325 biologically defined gene sets.

evaluates microarray data at the level of gene sets. The gene sets are defined based on prior biological knowledge, e.g., published information about biochemical pathways or coexpression in previous experiments. The goal of GSEA is to determine whether members of a gene set S tend to occur toward the top (or bottom) of the list L, in which case the gene set is correlated with the phenotypic class distinction.

We used a preliminary version of GSEA to analyze data from muscle biopsies from diabetics vs. healthy controls (4). The method revealed that genes involved in oxidative phosphorylation show reduced expression in diabetics, although the average decrease per gene is only 20%. The results from this study have been independently validated by other microarray studies (5) and by *in vivo* functional studies (6).

DAVID 功能富集分析

DAVID Bioinformatics Resources
Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

Home | Start Analysis | Shortcut to DAVID Tools | Technical Center | Downloads & APIs | Term of Service | About DAVID | About LHRI

Overview

The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID**) provides a comprehensive set of functional annotation tools for investigators to understand the biological meaning behind large lists of genes. These tools are powered by the comprehensive **DAVID Knowledgebase** built upon the DAVID Gene concept which pulls together multiple sources of functional annotations. For any given gene list, DAVID tools are able to:

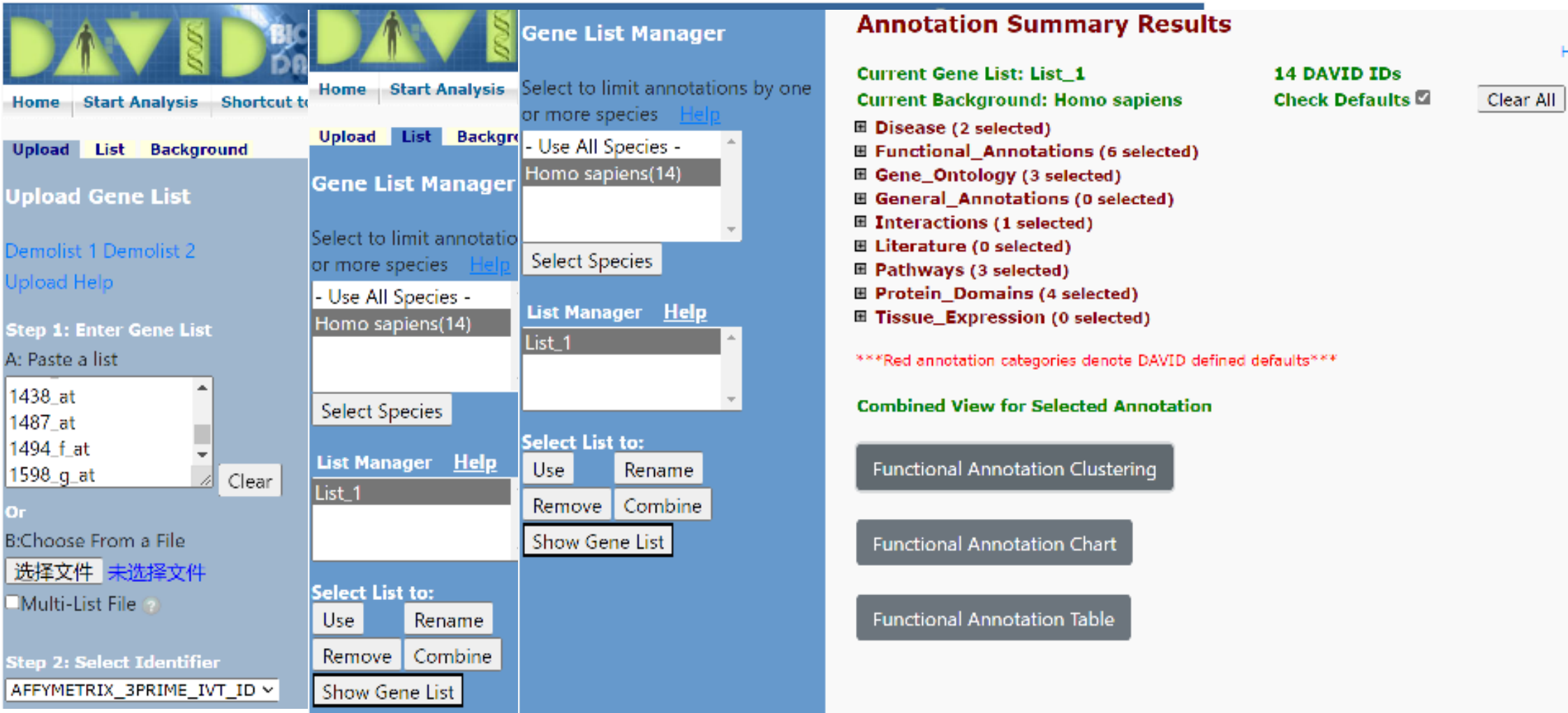
- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list
- List interacting proteins
- Explore gene names in batch

Hot Links

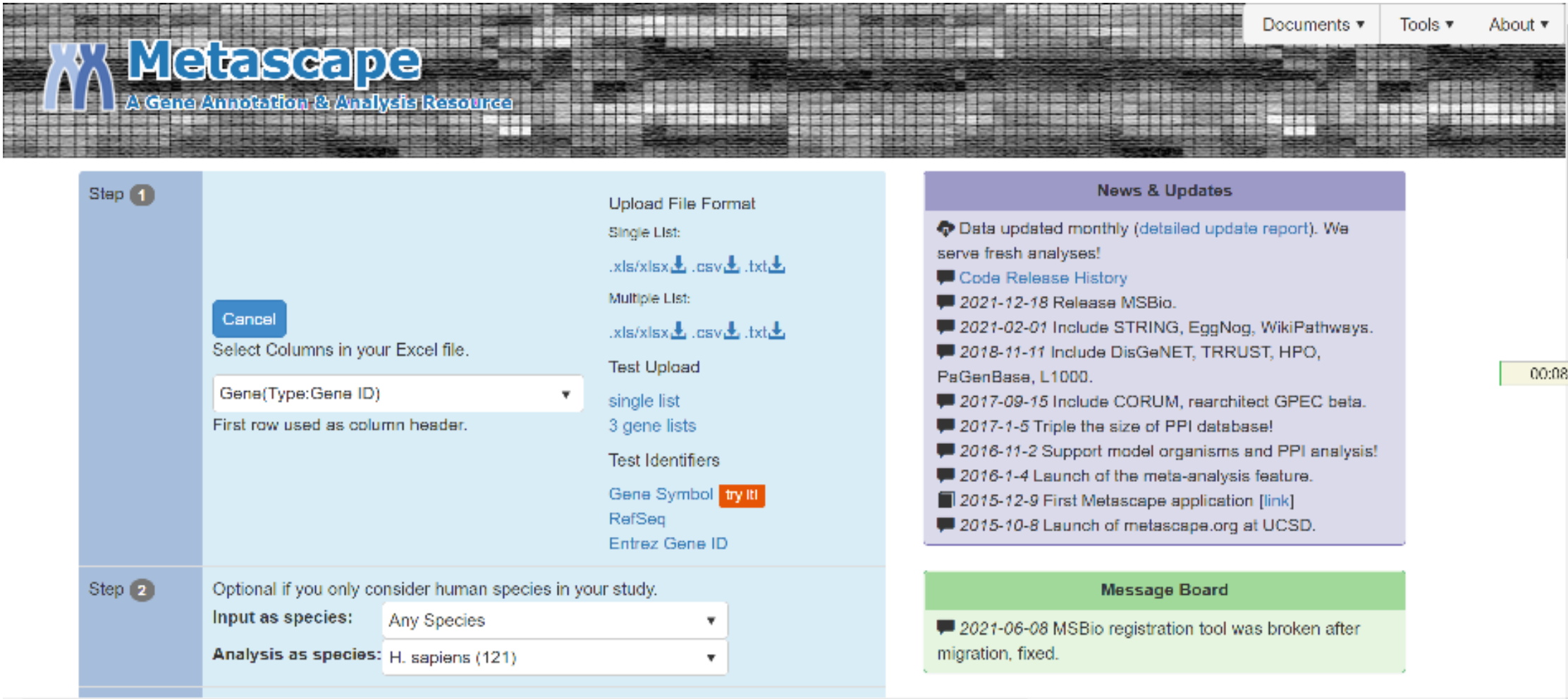
Multiple positions available in LHRI

The Laboratory of Human Retrovirology and Immunoinformatics (LHRI) has collaborated with the National Institute of Allergy and Infectious Diseases (NIAID) and supported NIAID clinical trials for patients infected with HIV mutants resisting anti-retroviral therapy. LHRI has isolated the multiple-class drug-resistant (MDR) variants from patients and characterized each variant's drug sensitivity and infectivity. The study aims to define salvage therapy and develop novel therapy (chemotherapy and immunotherapy). During the investigation, LHRI has characterized the emergence of novel mutations on drug susceptibility and viral replication. LHRI is a pioneer in researching the anti-viral cytokine, Interleukin-27, DNA-repair protein (Ku70)-mediated innate immune response against HIV and other virus co-infection, and novel subsets of immune cells. LHRI maintains the Database for Annotation, Visualization and Integrated Discovery (**DAVID**).

DAVID

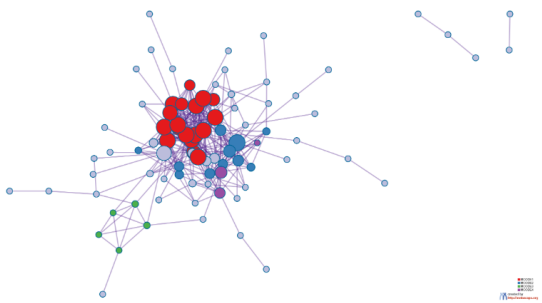

Home
Start Analysis
Shortcut t
Upload
List
Background
Upload Gene List
Demolist 1 Demolist 2
Upload Help
Step 1: Enter Gene List
A: Paste a list
1438_at
1487_at
1494_f_at
1598_g_at
Clear
Or
B:Choose From a File
选择文件
未选择文件
Multi-List File
Step 2: Select Identifier
AFFYMETRIX_3PRIME_IVT_ID
Home
Start Analysis
Upload
List
Gene List Manager
Select to limit annotatio
or more species
Help
- Use All Species -
Homo sapiens(14)
Select Species
List Manager
Help
List_1
Select List to:
Use
Rename
Remove
Combine
Show Gene List
Gene List Manager
Select to limit annotations by one
or more species
Help
- Use All Species -
Homo sapiens(14)
Select Species
List Manager
Help
List_1
Select List to:
Use
Rename
Remove
Combine
Show Gene List
Annotation Summary Results
Current Gene List: List_1
Current Background: Homo sapiens
14 DAVID IDs
Check Defaults
Clear All
Disease (2 selected)
Functional_Annotations (6 selected)
Gene_Ontology (3 selected)
General_Annotations (0 selected)
Interactions (1 selected)
Literature (0 selected)
Pathways (3 selected)
Protein_Domains (4 selected)
Tissue_Expression (0 selected)
Red annotation categories denote DAVID defined defaults
Combined View for Selected Annotation
Functional Annotation Clustering
Functional Annotation Chart
Functional Annotation Table

Metascape


Metascape
A Gene Annotation & Analysis Resource
Documents
Tools
About
Step 1
Cancel
Select Columns in your Excel file.
Gene(Type:Gene ID)
First row used as column header.
Upload File Format
Single List:
.xls/xlsx .csv .txt
Multiple List:
.xls/xlsx .csv .txt
Test Upload
single list
3 gene lists
Test Identifiers
Gene Symbol try it!
RefSeq
Entrez Gene ID
Step 2
Optional if you only consider human species in your study.
Input as species:
Any Species
Analysis as species:
H. sapiens (121)
News & Updates
Data updated monthly (detailed update report). We serve fresh analyses!
Code Release History
2021-12-18 Release MSBio.
2021-02-01 Include STRING, EggNog, WikiPathways.
2018-11-11 Include DisGeNET, TRRUST, HPO, PaGenBase, L1000.
2017-09-15 Include CORUM, rearchitect GPEC beta.
2017-1-5 Triple the size of PPI database!
2016-11-2 Support model organisms and PPI analysis!
2016-1-4 Launch of the meta-analysis feature.
2015-12-9 First Metascape application [link]
2015-10-8 Launch of metascape.org at UCSD.
Message Board
2021-06-08 MSBio registration tool was broken after migration, fixed.
00:08

Metascape

Figure 3. Protein-protein interaction network and MCODE components identified i



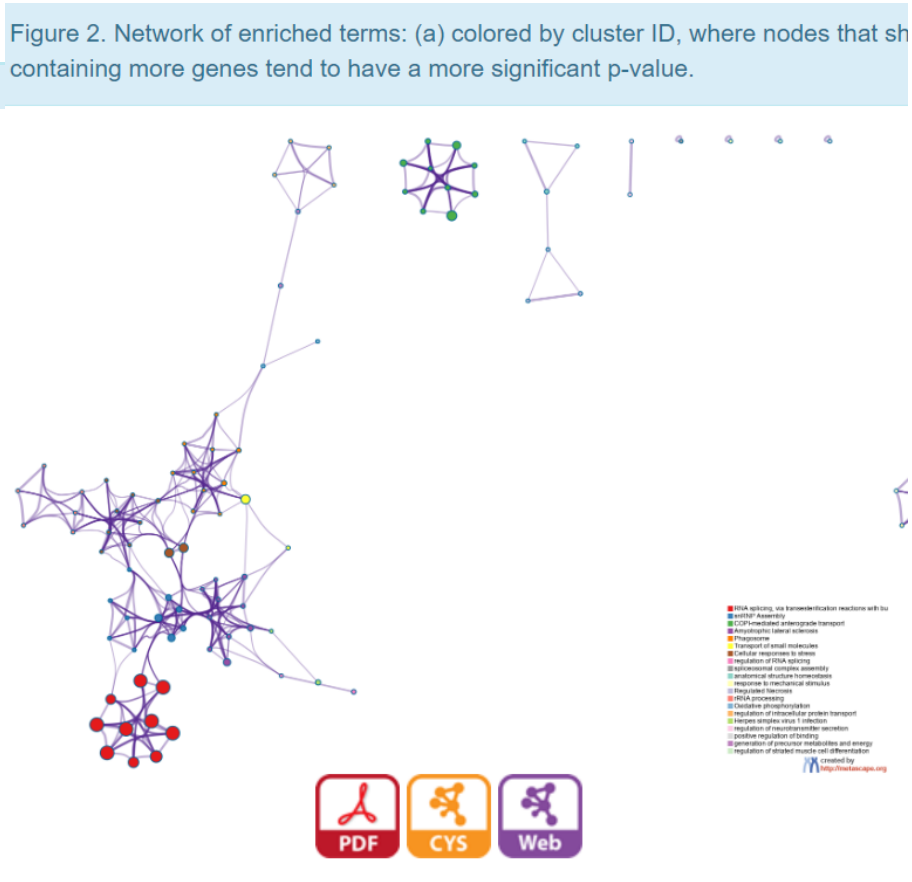


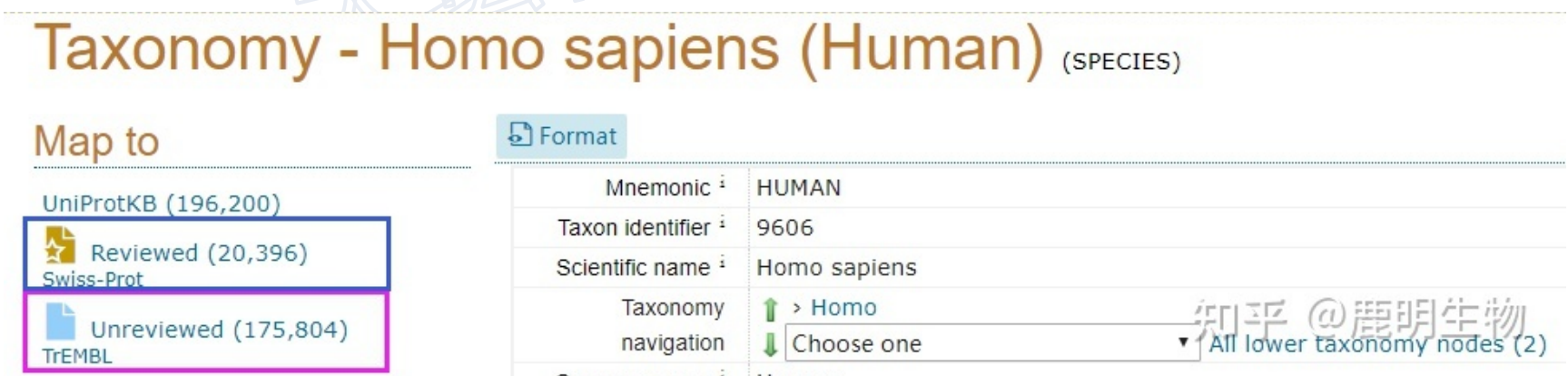
Figure 2. Network of enriched terms: (a) colored by cluster ID, where nodes that sh containing more genes tend to have a more significant p-value.

Zhou et al., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications (2019) 10(1):1523.

蛋白功能查询

- Uniprot
- NCBI
- Gene card

如果蛋白质组所研究的物种已经被测序，推荐使用Uniprot数据库作为搜库的数据库，通常我们用的是经过人工校验的蛋白数据，如果有特殊的研究目的想关注未注释的蛋白，或者研究的物种没有经过测序，校验信息非常少时，合并使用NCBI。**不是库越大越好，需要在全面性和准确性上做好平衡。**

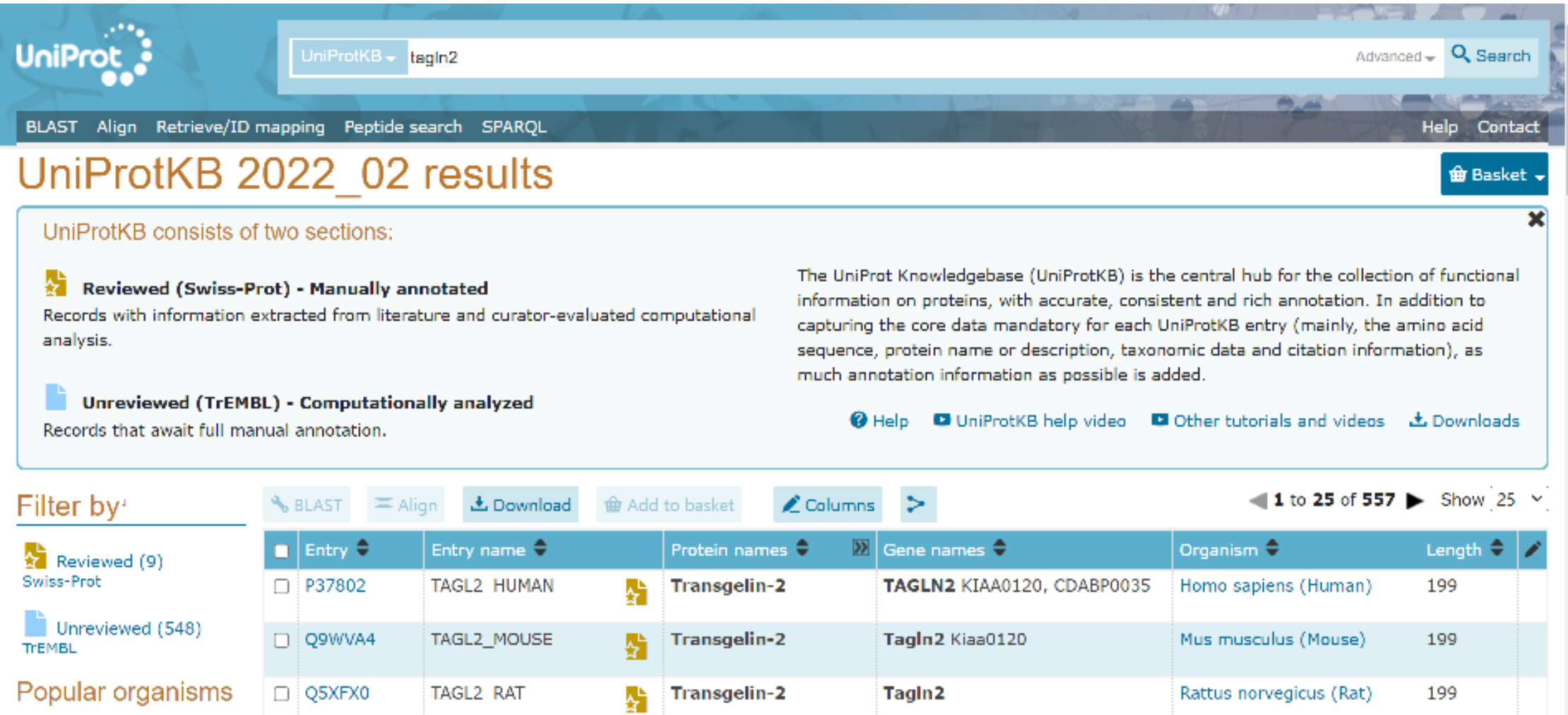


https://zhuanlan.zhihu.com/p/368209990

Uniprot

最全面、使用频率高、冗余度最低；网页打开缓慢

https://www.uniprot.org/



UniProtKB - P37802 (TAGL2_HUMAN)

Basket

Display | Help video | BLAST | Align | Format | Add to basket | History | Add a publication | Feedback

- Entry
- Publications
- Feature viewer
- Feature table

None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequences (2+)

Protein	**Transgelin-2**
Gene	**TAGLN2**
Organism	*Homo sapiens (Human)*
Status	Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level

Function

GO - Molecular function

- cadherin binding — Source: BHF-UCL

Complete GO annotation on QuickGO ...

GO - Biological process

- epithelial cell differentiation — Source: UniProtKB

Complete GO annotation on QuickGO ...

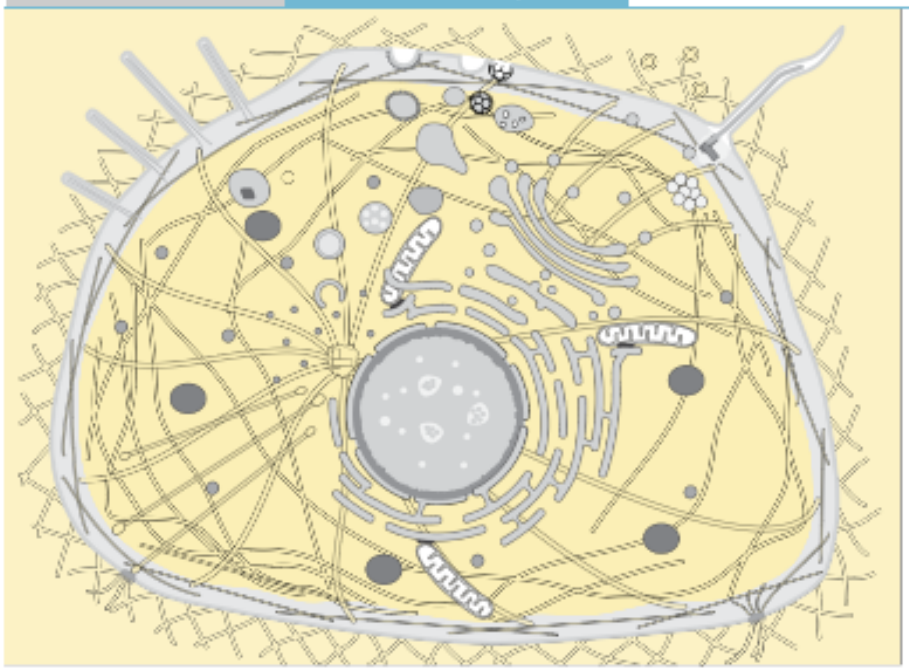
Enzyme and pathway databases

PathwayCommons	P37802
Reactome	R-HSA-114608, Platelet degranulation
SignaLink	P37802

MIM	604634, gene
neXtProt	NX_P37802
VEuPathDB	HostDB:ENSG00000158710

Subcellular location

UniProt annotation | GO Cellular component



Automatic annotation
Manual annotation
Source: SwissBioPics

Cytoskeleton
- cytoskeleton — Source: GO_Central

Cytosol
- cytosol — Source: Reactome

Extracellular region or secreted
- extracellular exosome — Source: UniProtKB
- extracellular region — Source: Reactome

Other locations
- vesicle — Source: UniProtKB

Complete GO annotation on QuickGO ...

Interaction

Binary interactions

P37802 has binary interactions with 5 proteins

Filter

Subcellular location: Select...

Diseases: Select...

Clear

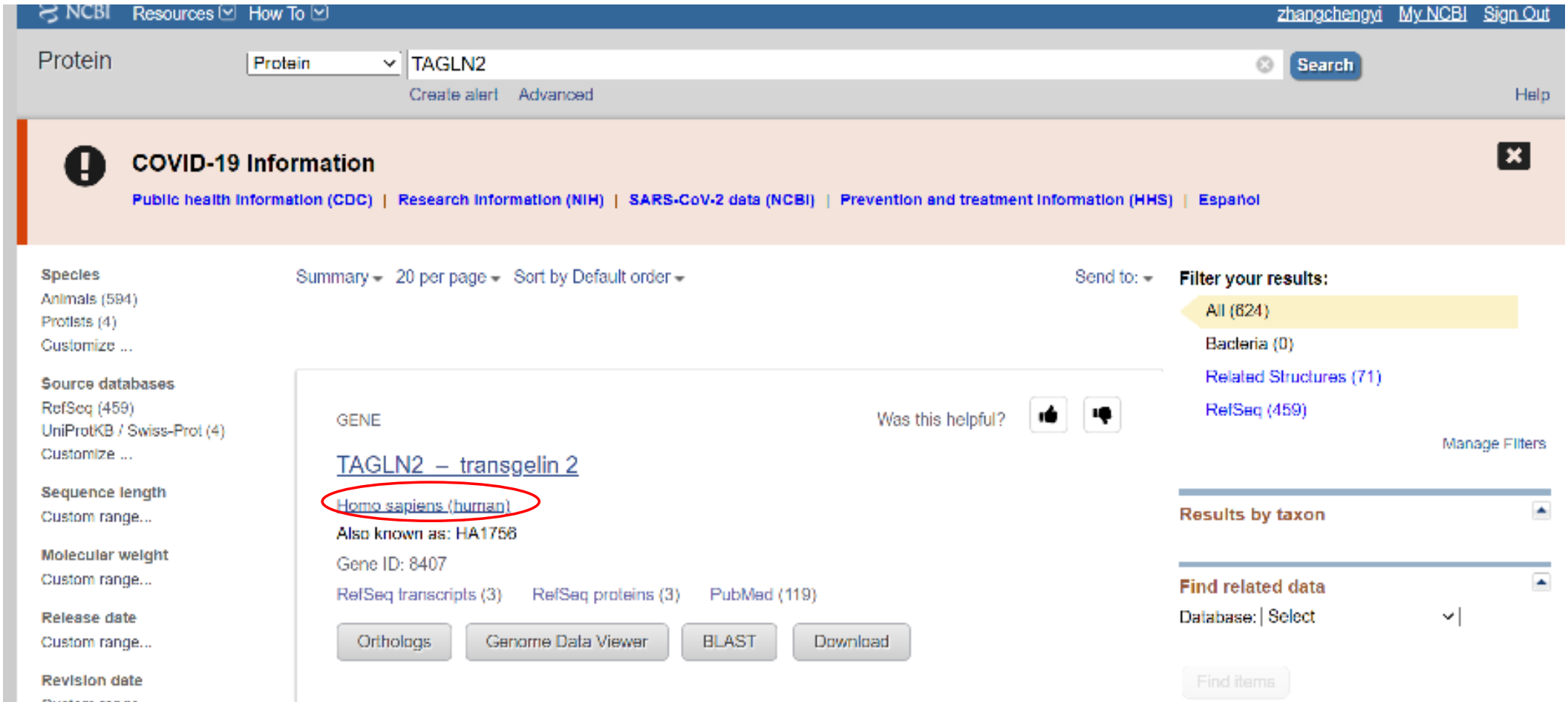
Hide details

P37802

With	#Exp.	IntAct
ACTB [P60709]	3	EBI-1056740,EBI-353944
DNM2 - isoform 2 [P50570-2]	3	EBI-1056740,EBI-10968534
GDAP1 [Q8TB36]	3	EBI-1056740,EBI-11110431

NCBI 最全面最权威

www.ncbi.nlm.nih.gov



TAGLN2 transgelin 2 [*Homo sapiens* (human)]

Download Datasets

Gene ID: 8407, updated on 13-Feb-2022

Summary

Official Symbol	TAGLN2 provided by HGNC
Official Full Name	transgelin 2 provided by HGNC
Primary source	HGNC:HGNC:11554
See related	Ensembl:ENSG00000158710 MIM:604634; AllianceGenome:HGNC:11554
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	HA1756
Summary	The protein encoded by this gene is similar to the protein transgelin, which is one of the earliest markers of differentiated smooth muscle. The specific function of this protein has not yet been determined, although it is thought to be a tumor suppressor. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Mar 2013]
Expression	Ubiquitous expression in lung (RPKM 268.8), stomach (RPKM 231.0) and 25 other tissues See more
Orthologs	mouse all

NEW **Try the new Gene table**

Try the new Transcript table

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Expression
- Bibliography
- Phenotypes
- Variation
- HIV-1 interactions
- Interactions
- General gene information
  - Markers, Related pseudogene(s), Potential readthrough, Clone Names, Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links

Bibliography

Related articles in PubMed 相关文献

1. Transgelin-2 in Multiple Myeloma: A New Marker of Renal Impairment?
   Woziwodzka K, *et al*. Molecules, 2021 Dec 23. PMID 35011306, **Free PMC Article**
2. Downregulation of transgelin 2 promotes breast cancer metastasis by activating the reactive oxygen species/nuclear factor-κB signaling pathway.
   Yang L, *et al*. Mol Med Rep, 2019 Nov. PMID 31485630, **Free PMC Article**
3. Transgelin-2: Biochemical and Clinical Implications in Cancer and Asthma.
   Yin LM, *et al*. Trends Biochem Sci, 2019 Oct. PMID 31256982, **Free PMC Article**
4. Transgelin-2 expression in breast cancer and its relationships with clinicopathological features and patient outcome.
   Hao R, *et al*. Breast Cancer, 2019 Nov. PMID 31144206
5. Transgelin 2 overexpression inhibits cervical cancer cell invasion and migration.
   Zhou Q, *et al*. Mol Med Rep, 2019 Jun. PMID 30942422

See all (119) citations in PubMed

See citations in PubMed for homologs of this gene provided by HomoloGene

GeneRIFs: Gene References Into Functions

What's a GeneRIF?

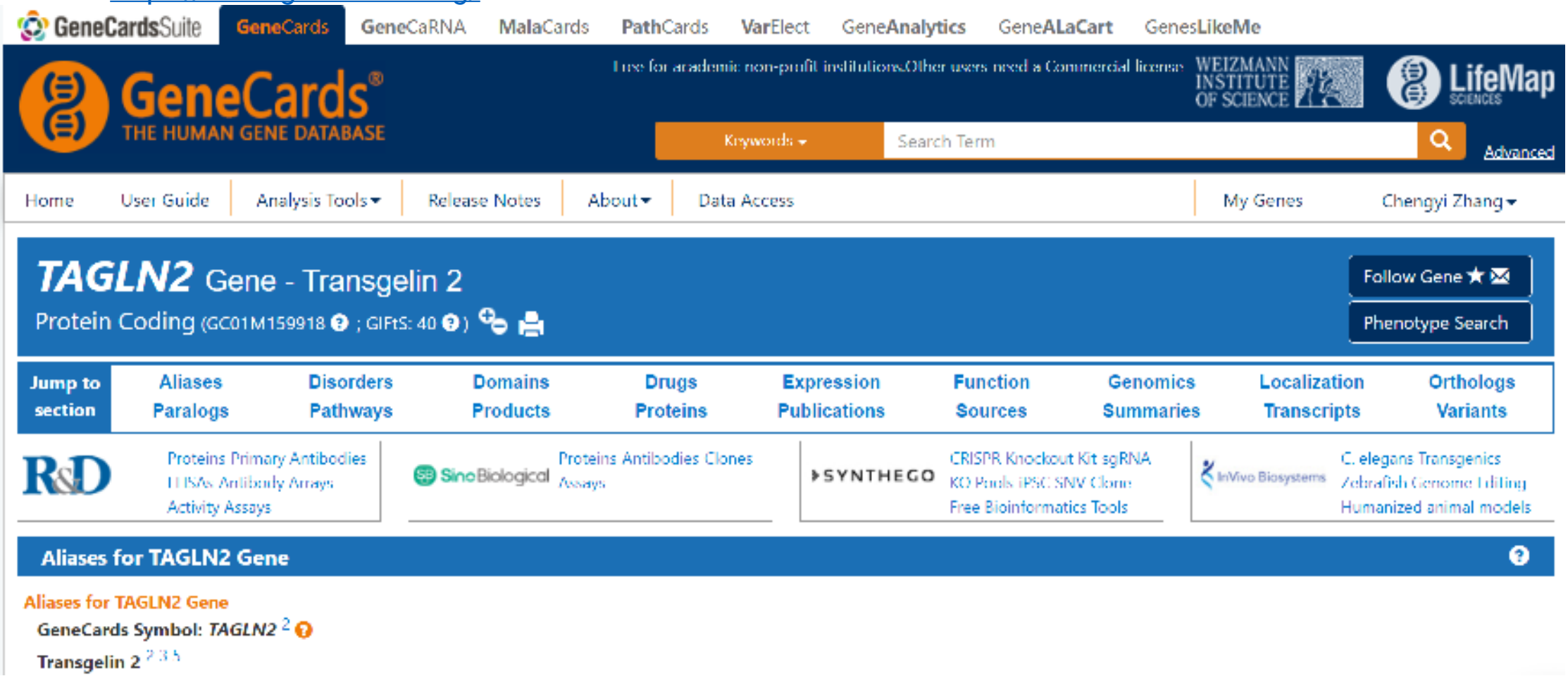
功能相关文献

1. Transgelin-2 in Multiple Myeloma: A New Marker of Renal Impairment?
2. Transgelin-2 and phosphoregulation of the LIC2 subunit of dynein govern mitotic spindle orientation.
3. Transgelin-2 contributes to proliferation and progression of hepatocellular carcinoma via regulating Annexin A2.
4. REVIEW: Biochemical and Clinical Implications in Cancer and Asthma
5. The present study proposed TAGLN2 to function as a tumor suppressor and that loss of TAGLN2 may promote the metastasis of breast cancer by activating the ROS/NFkappaB signaling pathway.
6. Transgelin-2 was highly overexpressed in breast cancer and relevant to progression. High transgelin-2 expression might predict poor outcome in patients with ER-negative tumors.
7. dual functional nature of TAGLN2-G-actin polymerization and Arp2/3 complex inhibition-may account for the mechanisms of filopodia development at the edge of Arp2/3-rich lamellipodia in various cell types
8. TAGLN 2 overexpression in HeLa cells could inhibit cell viability, migration and invasion, and it was suggested that this may occur via upregulation of the expression levels of E cadherin and inhibitor of nuclear factor kappalightchainenhancer of activated B cells (NFkappaB) (IkappaB), and downregulation of C XC chemokine receptor type 4, matrix metalloproteinase (MMP)2, MMP9, p50 and transcription factor p65.
9. Authors found transgelin-2 expression was induced by KRAS mutation. In the case of KRAS mutation, ERK2 interacted with 29-31 amino acids of transgelin-2 and subsequently phosphorylated the S145 residue of transgelin-2.
10. Study provides important evidence that hypoxia-inducible TAGLN2 is involved in the selection of cancer cells with enhanced EMT properties to overcome the detrimental environment of cancer cells as gamma radiation.

GeneCards

基因百科数据库，打开稍慢

https://www.genecards.org/



Summaries for TAGLN2 Gene

Entrez Gene Summary for TAGLN2 Gene

The protein encoded by this gene is similar to the protein transgelin, which is one of the earliest markers of differentiated smooth muscle. The specific function of this protein has not yet been determined, although it is thought to be a tumor suppressor. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Mar 2013]

GeneCards Summary for TAGLN2 Gene

TAGLN2 (Transgelin 2) is a Protein Coding gene. Diseases associated with TAGLN2 include Barrett's Adenocarcinoma and Maxillary Sinus Cancer. Among its related pathways are Response to elevated platelet cytosolic Ca2+. Gene Ontology (GO) annotations related to this gene include *actin filament binding*. An important paralog of this gene is TAGLN3.

Gene Wiki entry for TAGLN2 Gene

Additional gene information for TAGLN2 Gene

HGNC (11554) NCBI Entrez Gene (8407) Ensembl (ENSG00000158710) OMIM® (604634) UniProtKB/Swiss-Prot (P37802) Open Targets Platform(ENSG00000158710)

Alliance of Genome Resources

Drugs & Compounds for TAGLN2 Gene

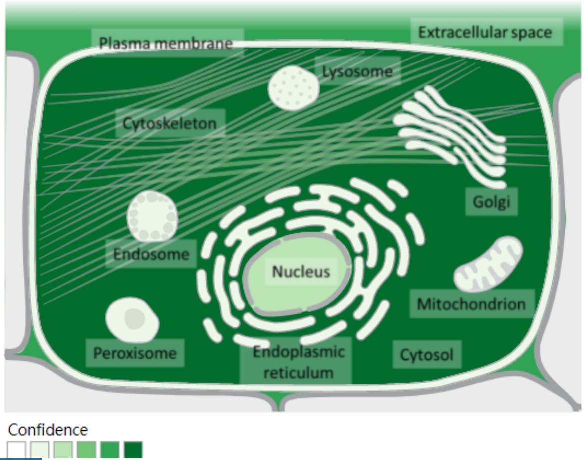
Products: Drug products for research

(1) Drugs for TAGLN2 Gene - From: DrugBank

Filter: *(1 result)* Options

	Name	Status	Disease Links	Group	Role	Mechanism of Action	Clinical Trials
⊕	Artenimol [23]	Approved, Experimental, Investigational [23]	MalaCards Medline Plus	Pharma	Target, ligand		

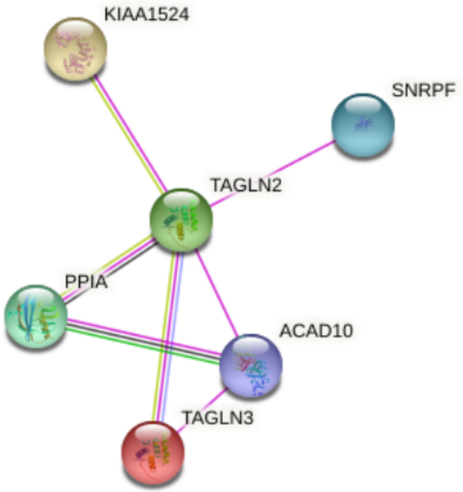
Subcellular locations from COMPARTMENTS



Compartment	Confidence
cytosol	5
extracellular	4
cytoskeleton	3
nucleus	2
plasma membrane	1
mitochondrion	1
peroxisome	1
endoplasmic reticulum	1
endosome	1
lysosome	1

Interacting Proteins for TAGLN2 Gene

STRING Interaction Network Preview (showing top 5 STRING interactants - click image to see top 25)



GeneAnalytics

实验试剂订购

Search for latest publications for TAGLN2 gene in PubMed and other databases

Products for TAGLN2 Gene

R&D SYSTEMS
a bio-techne brand

› Learn more about R&D Systems custom TAGLN2 antibody, protein, and immunoassay development services

ORIGENE
Your Gene Company

› OriGene Antibodies for TAGLN2 See all 11 »
TA812766 TA812766AM
Custom Antibody Services
› Browse OriGene ELISA Kits
› Custom Assay Services
› OriGene Purified Proteins for TAGLN2 See all 9 »
AR09878PU-L AR09878PU-N
› Search Origene for MassSpec and Protein Over-expression Lysates for TAGLN2
› Origene Custom Protein Services for TAGLN2
› Origene sirna-oligo-duplexes, shrna-plasmids, and shrna-lentiviral-particles products in human, mouse, rat for TAGLN2
› Browse OriGene Inhibitory RNA Products for TAGLN2
› OriGene Primers for TAGLN2 See all 4 »
HK207706 HP207061
› OriGene CRISPR knockouts for TAGLN2 See all 4 »
GA105554 GA704701
› OriGene clones in human,mouse,rat for TAGLN2 See all 6 »
MC200788 RN704860
› OriGene ORF clones in human,mouse,rat for TAGLN2 See all 25 »
MG202012 MR202012
› Custom cloning services - gene synthesis, subcloning, mutagenesis, variant library, vector

Publications for TAGLN2 Gene

Filter: (148 results) See all 148 »

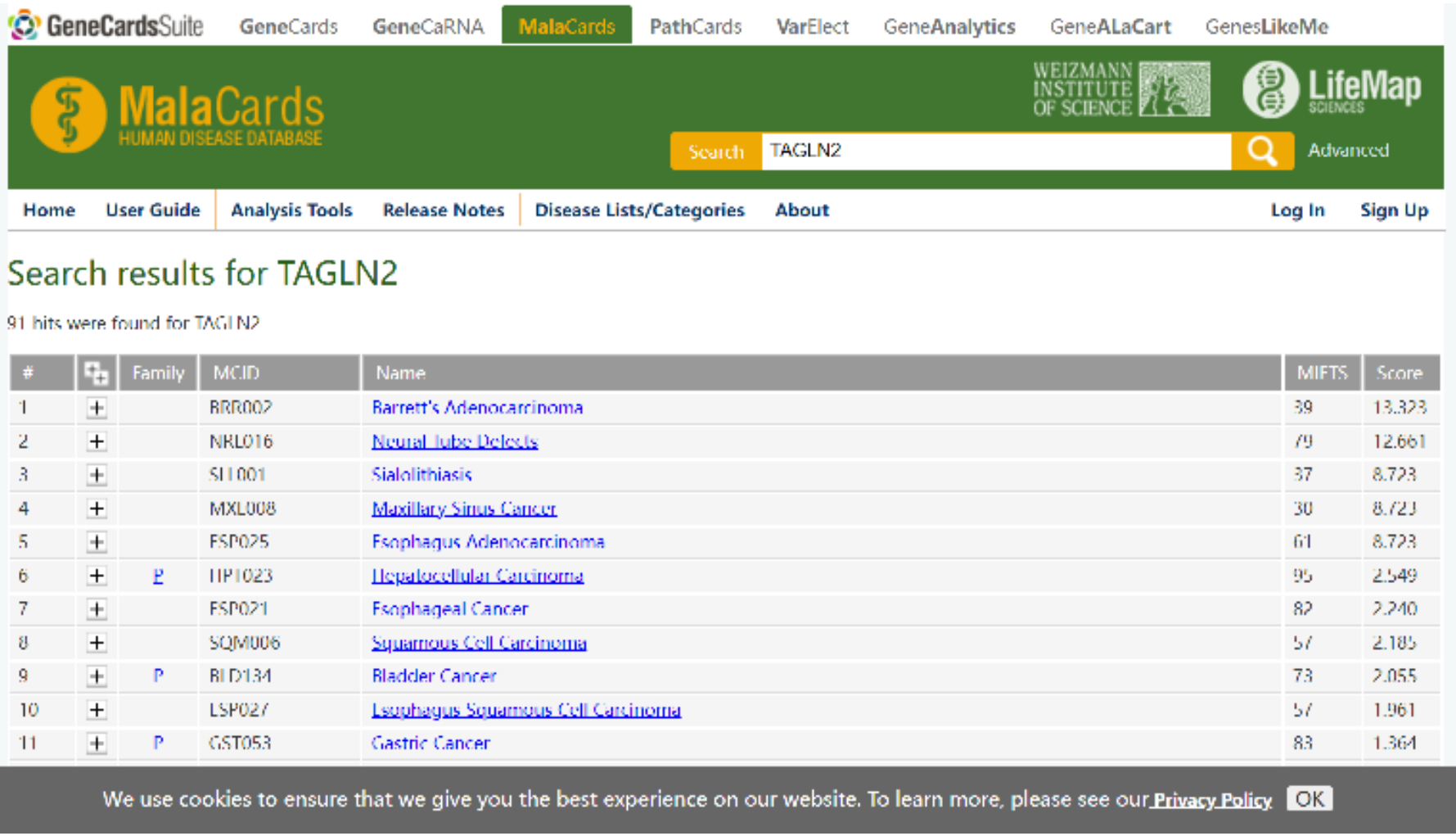
1. Systematic mapping and functional analysis of a family of human epididymal secretory sperm-located proteins. (PMID: 20736409) Li J ... Liu Y *Molecular & cellular proteomics : MCP 2010* [3] [4]
2. Tissue proteomics reveals differential and compartment-specific expression of the homologs transgelin and transgelin-2 in lung adenocarcinoma and its stroma. (PMID: 19848416) Rho JH ... Wang JY *Journal of proteome research 2009* [3] [21]
3. The DNA sequence and biological annotation of human chromosome 1. (PMID: 16710414) Gregory SG ... Prigmore E *Nature 2006* [3] [4]
4. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). (PMID: 15489334) Gerhard DS ... MGC Project Team *Genome research 2004* [3] [4]
5. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. (PMID: 12665801) Gevaert K ... Vandekerckhove J *Nature biotechnology 2003* [3] [4]

Search for latest publications for TAGLN2 gene in PubMed and other databases

其他相关数据库

- 疾病相关数据库——OMIM、TCGA等
- 基因表达数据库——Expression Atlas - EMBL-EBI、GTEx、BodyMap、Cancer RNA-Seq Nexus等
- 基因功能数据库——Reactome等
- DNA变异数据库——dbSNP、dbVar、VarScan等
- 蛋白相互作用数据库——String、HPRD、BioGRID等
- 转录因子调节关系数据库——Cistrome、TRRD、TRANSFAC等

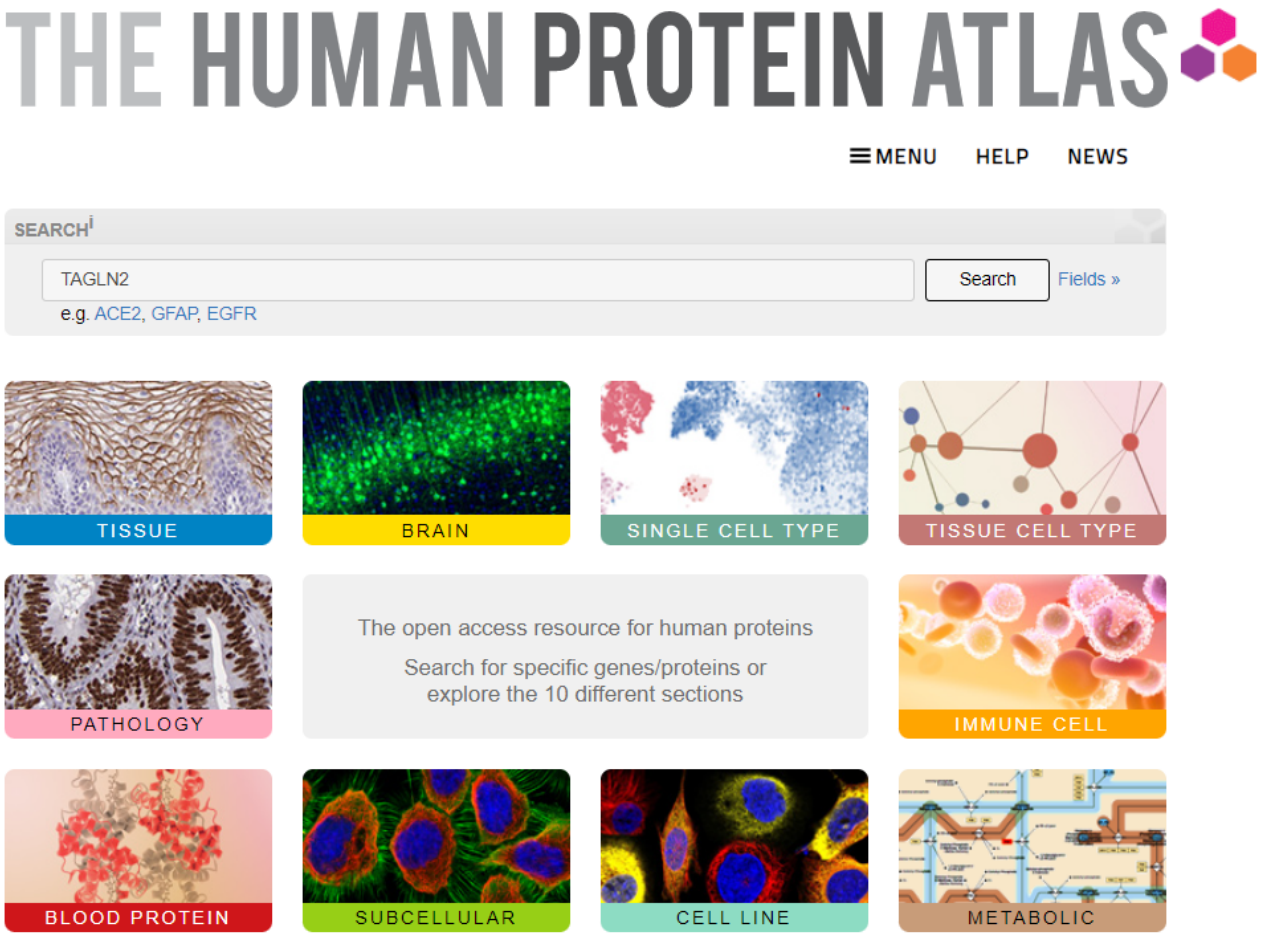
https://www.malacards.org/ 基因相关疾病查询

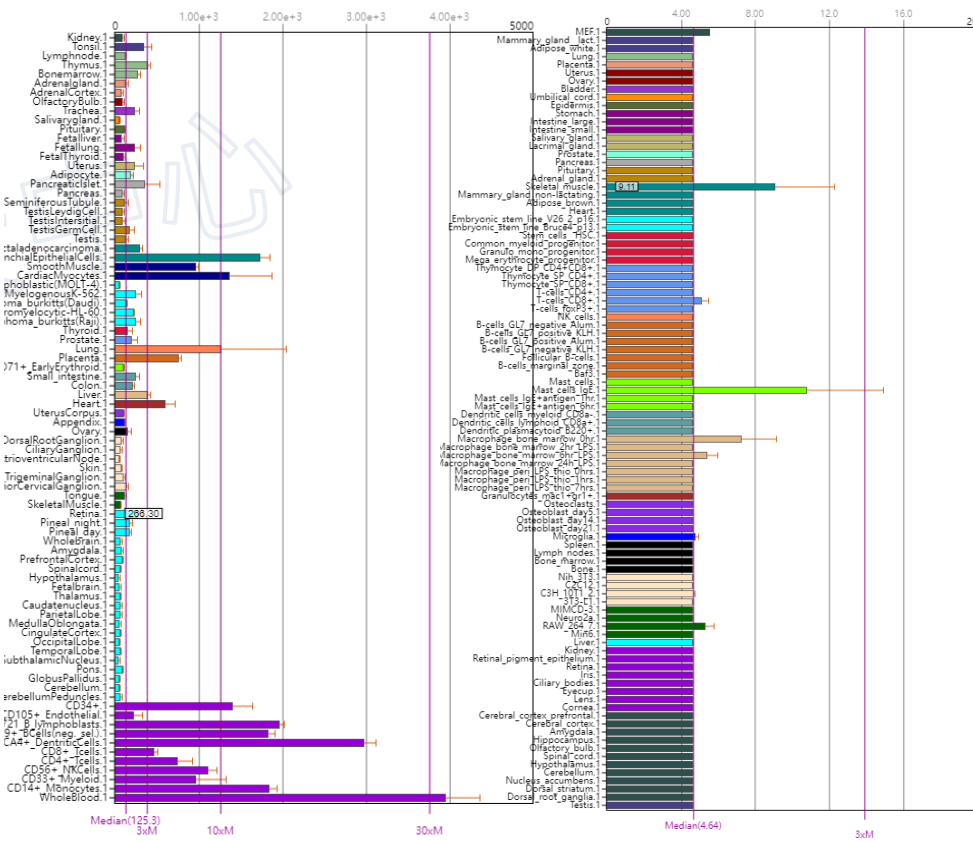


基因表达

https://www.proteinatlas.org/

http://biogps.org/



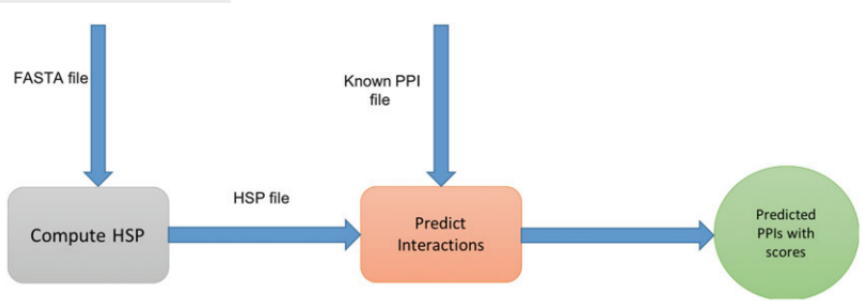


https://reactome.org/ 基因功能

蛋白间相互作用

A non-exhaustive list of commonly used molecular interaction networks

Network	Reference	Description	Species
BioGrid	[12]	Curated resource integrating protein, genetic, and phosphorylation interaction data from publications	13 major model orga man
GeneMania	[13]	Uses an algorithm to determine association strength based on publicly available data for protein and genetic interactions, pathways, co-expression, co-localization, and protein domain similarity	9
HTRIdb	[14]	Network of interactions between transcription factors and target genes	Human
HPRD	[15]	The first comprehensive human protein–protein interaction network	Human
IntAct	[16]	Curated resource for all types of molecular interactions	Many
I2D	[17]	Integration of known and predicted protein–protein interactions	6
STRING	[18]	Protein–protein interactions, experimental and predicted	>2000



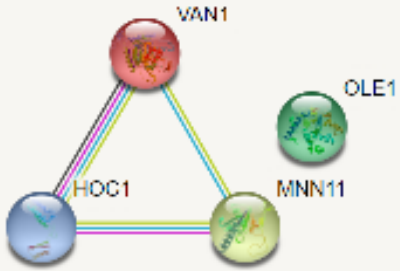
STRING 蛋白互作查询+功能富集分析

Version: 11.5 LOGIN | REGISTER | SURVEY

Version: 11.5 LOGIN | REGISTER | SURVEY

STRING

Search Download Help My Data



Viewers Legend Settings Analysis Exports Clusters More Less

Nodes:

Network nodes represent proteins

splice isoforms or post translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Color

colored nodes: query proteins and first shell of interactors

white nodes: second shell of interactors

Node Content

empty nodes: proteins of unknown 3D structure

filled nodes: some 3D structure is known or predicted

Edges: