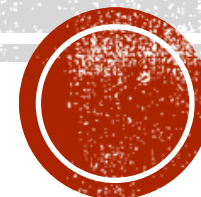


# GO, KEGG и все-все-все



Анастасия Жарикова, Юлия Алешина

5 апреля 2024 – ФББ МГУ

# ДАНО

- Набор интересных белков или генов

# ДАНО

- Набор интересных белков или генов
- Откуда такой набор можно взять?

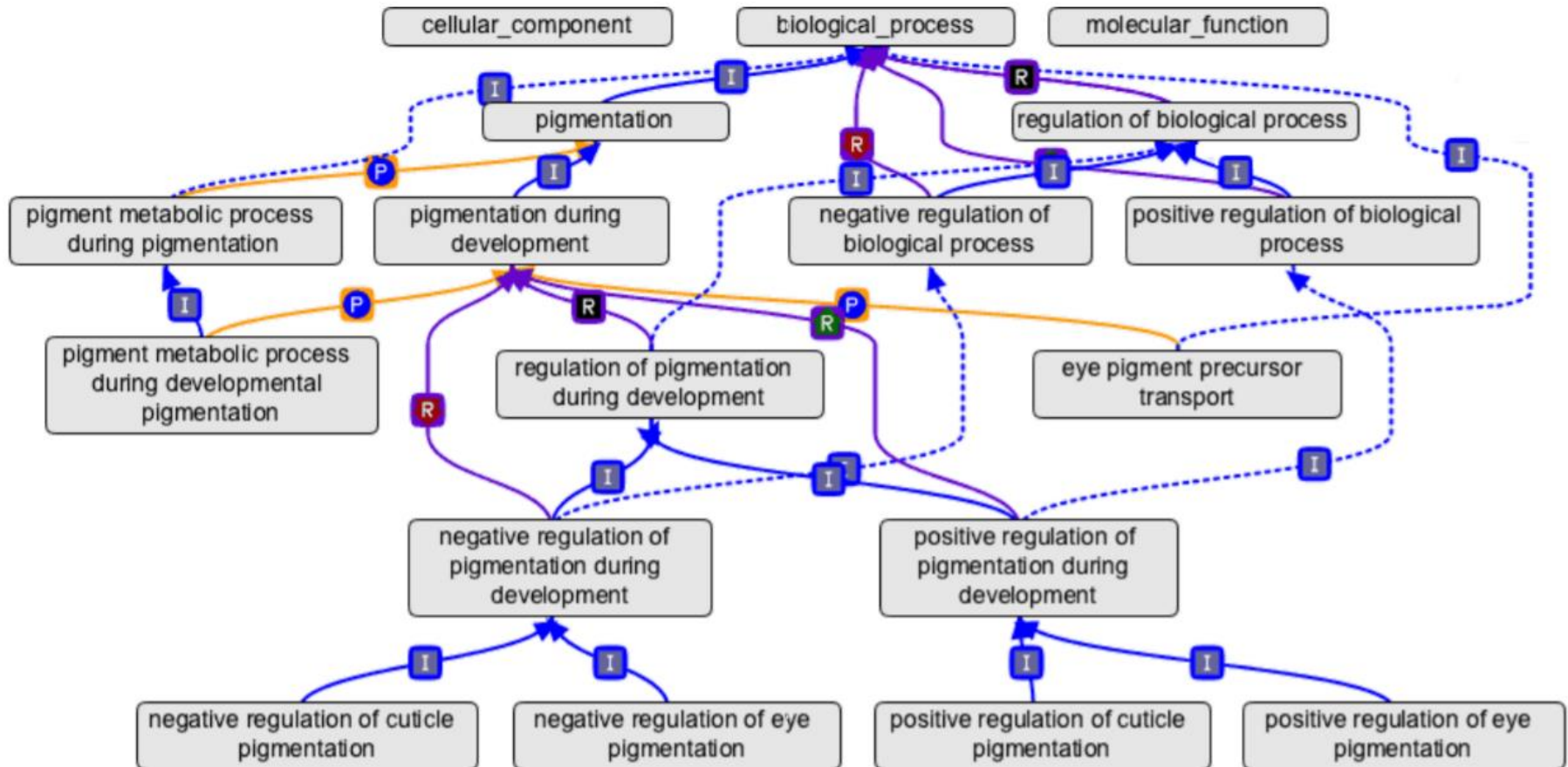
# ДАНО

- Набор интересных белков или генов
- Откуда такой набор можно взять?
  
- Экспрессионный профиль
- Дифференциальная экспрессия
- Масспектрометрия
- Данные из статей
- Любимый набор
- ...

Что дальше делать?

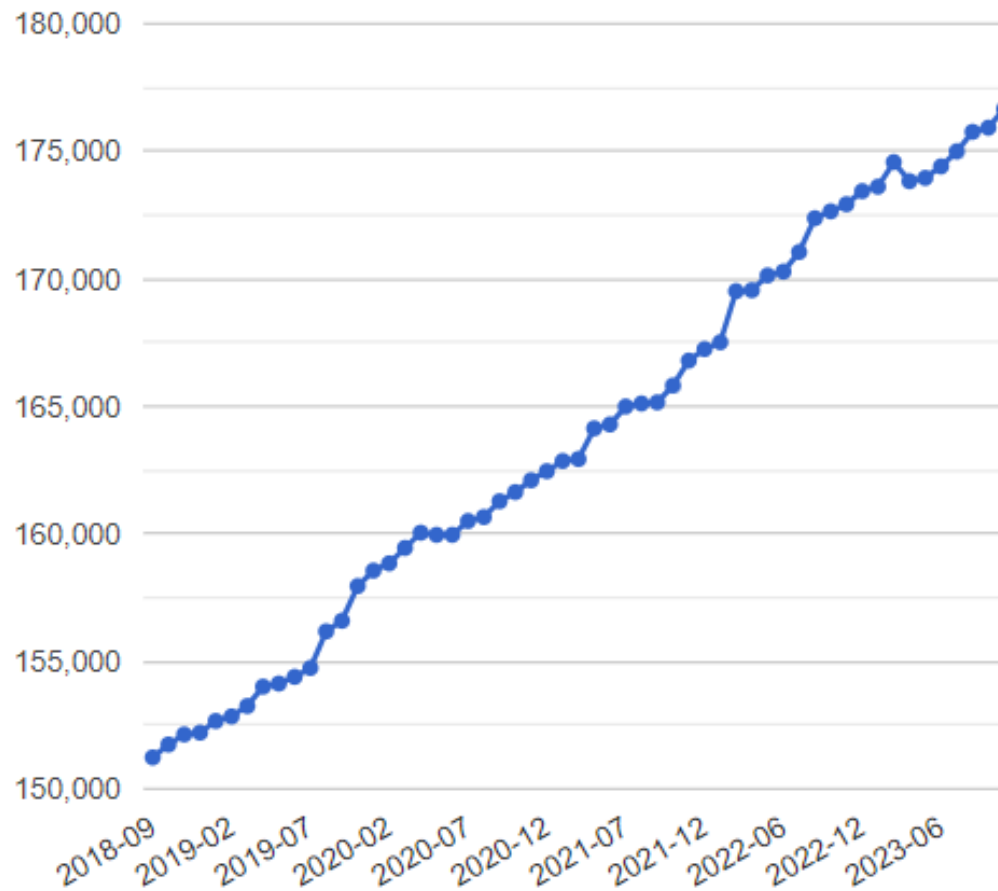
# GENE ONTOLOGY – GO

- <http://geneontology.org/>
- База данных GO представляет собой граф биологических терминов, соединенных различными отношениями





# КОЛИЧЕСТВО ПУБЛИКАЦИЙ, В КОТОРЫХ ИСПОЛЬЗУЕТСЯ GO







## Ontology

Property	Value
Valid terms	42442 ( $\Delta = -327$ )
Obsoleted terms	5287 ( $\Delta = 397$ )
Merged terms	2440 ( $\Delta = 0$ )
Biological process terms	27186
Molecular function terms	11198
Cellular component terms	4058

## Annotations

Property	Value
Number of annotations	7,655,937
Annotations for biological process	2,833,885
Annotations for molecular function	2,459,729
Annotations for cellular component	2,362,323
Annotations for evidence PHYLO	3,819,045
Annotations for evidence IEA	1,714,824
Annotations for evidence OTHER	858,807
Annotations for evidence EXP	955,756
Annotations for evidence ND	248,947
Annotations for evidence HTP	58,558
Number of annotated scientific publications	176,659

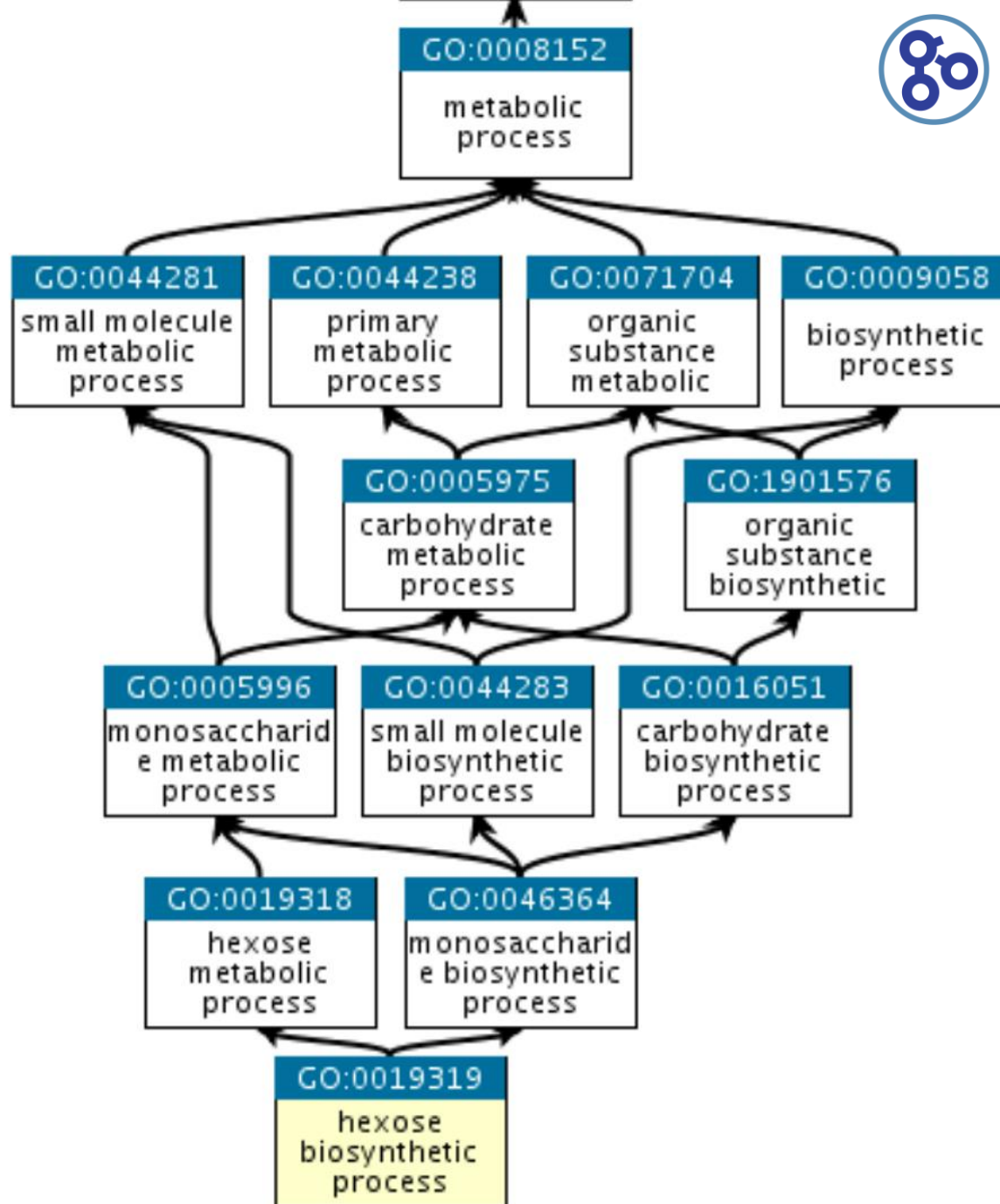
## Gene products and species

Property	Value
Annotated gene products	1,537,348
Annotated species	5,387
Annotated species with over 1,000 annotations	188

**Молекулярные функции** (англ. molecular function) — специфическая активность генного продукта на молекулярном уровне, например, связывание углеводов или АТФазная активность.

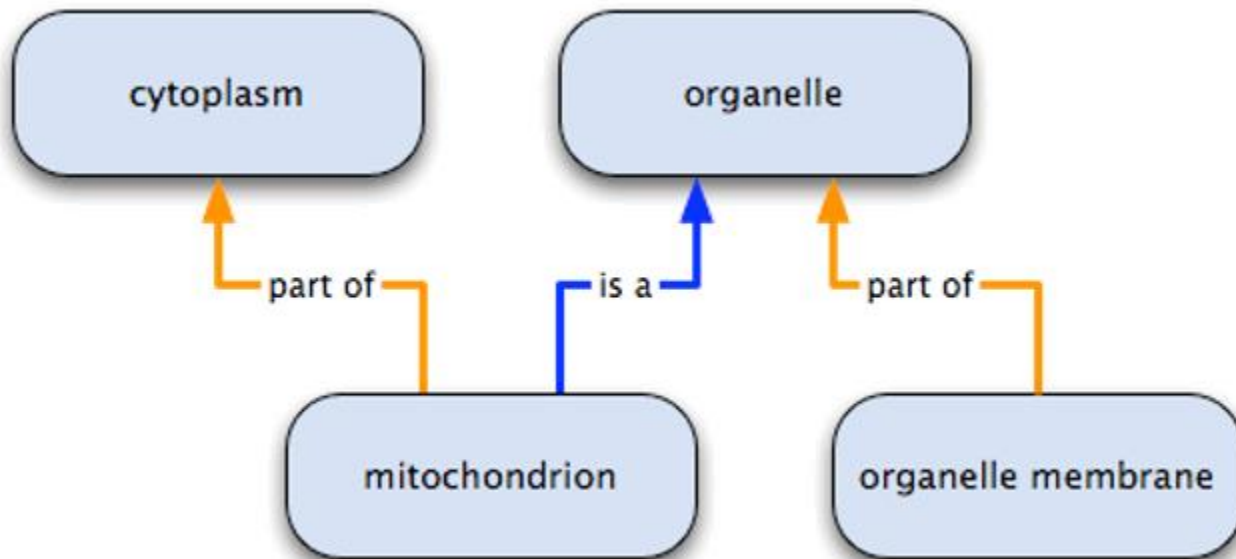
**Биологические процессы** (англ. biological process) — сложные явления, необходимые для жизнедеятельности организмов и происходящие благодаря осуществлению последовательности молекулярных функций, например, митоз или биосинтез пуринов.

**Клеточные компоненты** (англ. cellular component) — части клетки или внеклеточного пространства, где осуществляется функция генного продукта, например, ядро или рибосома.



Термин GO (GO term) – узел графа – может иметь любое количество связей с любыми другими узлами

Отношения между узлами – ребра графа бывают разного типа

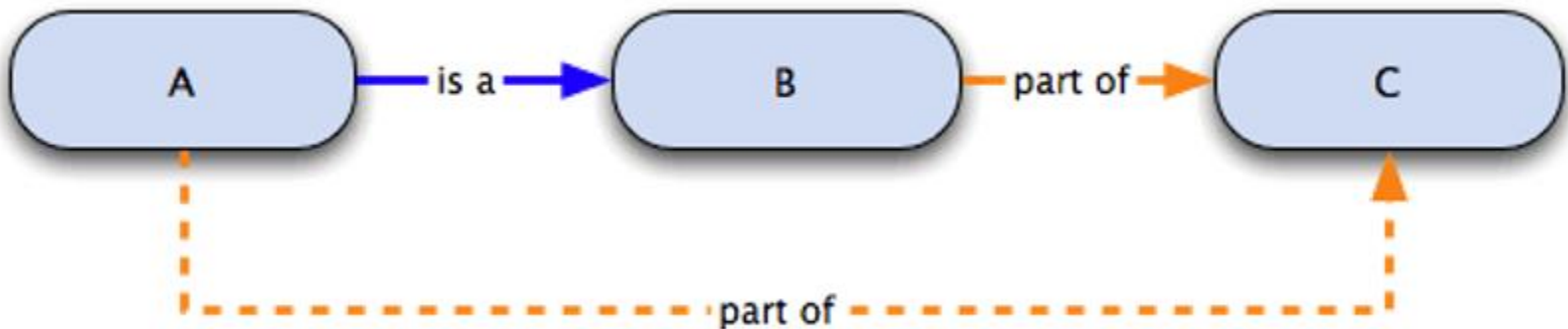


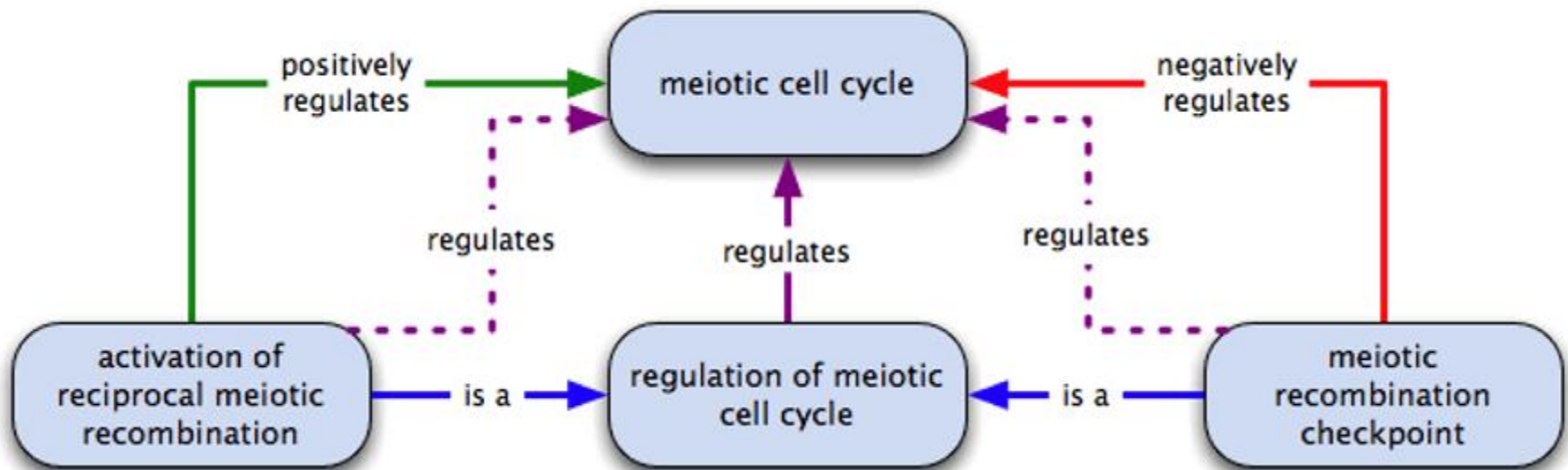
A *is a* B

B *part of* C

Значит, A *part of* C

→ Указанные  
→ непосредственно  
- - - - - Логически вытекающее





Inferred from Electronic Annotation (**IEA**) →

Inferred from Sequence or structural Similarity (**ISS**)

- Inferred from Sequence Orthology (**ISO**)
- Inferred from Sequence Alignment (**ISA**)
- Inferred from Sequence Model (**ISM**)

Inferred from Experiment (**EXP**)

- Inferred from Direct Assay (**IDA**)
- Inferred from Physical Interaction (**IPI**)
- Inferred from Mutant Phenotype (**IMP**)
- Inferred from Genetic Interaction (**IGI**)
- Inferred from Expression Pattern (**IEP**)

Traceable Author Statement (**TAS**)

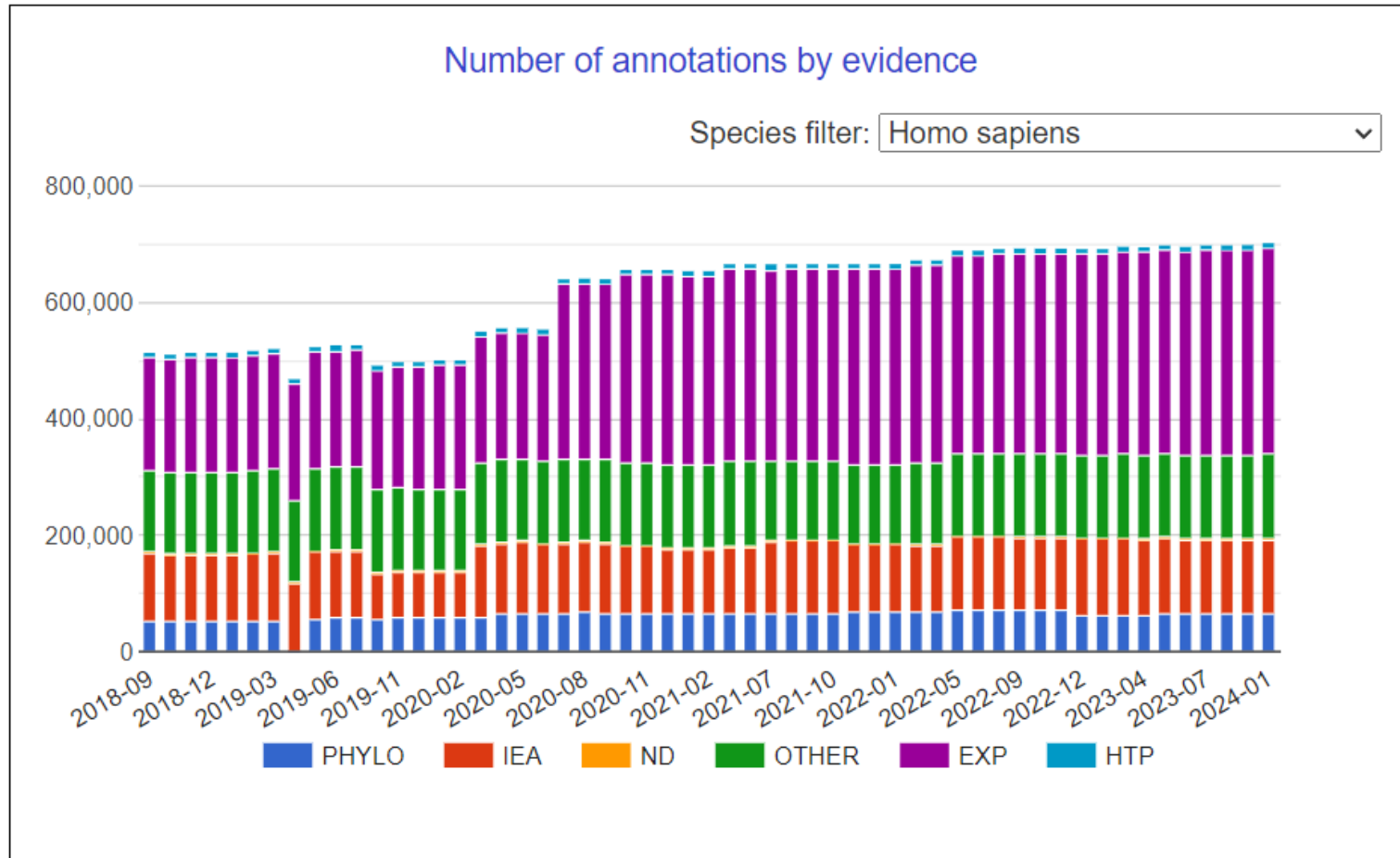
Non-traceable Author Statement (**NAS**)





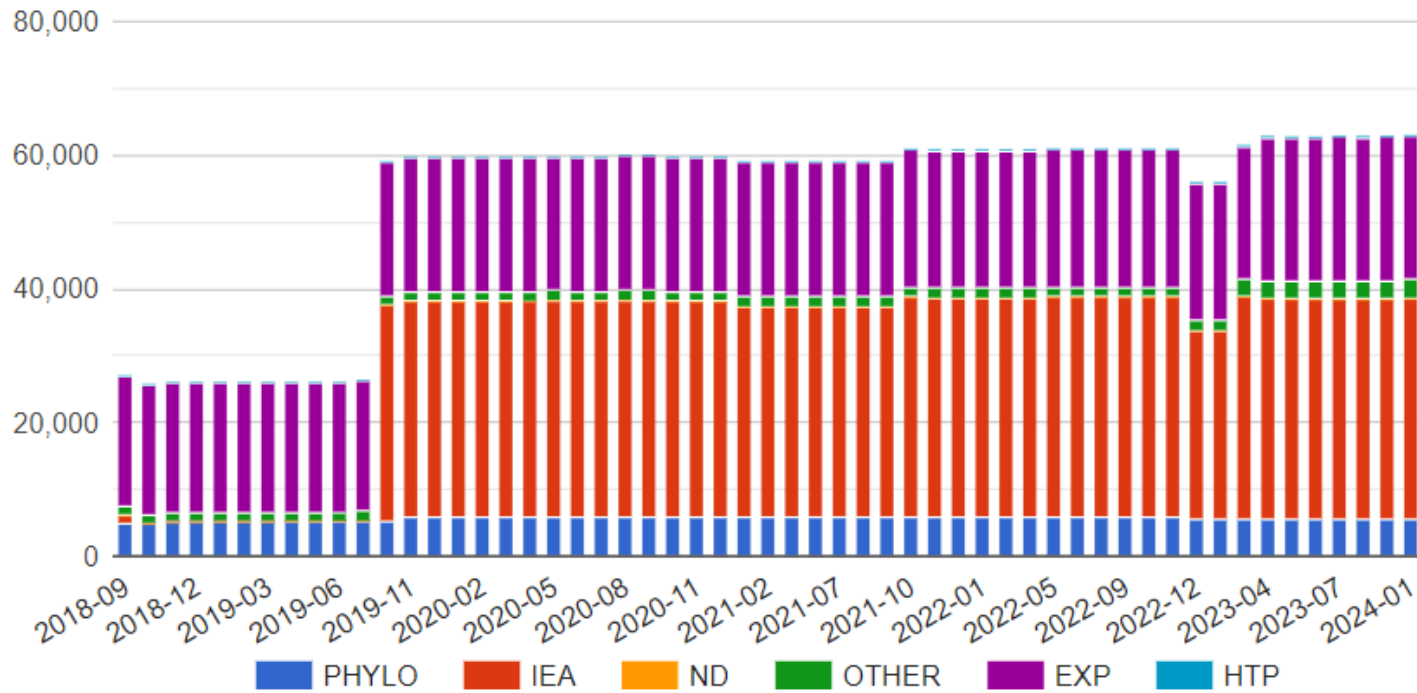
<input type="checkbox"/> Gene/product	Gene/product name	Annotation qualifier	GO class (direct)	Annotation extension	Contributor	Organism	Evidence	Evidence with	PANTHER family	Type	Isoform
<input type="checkbox"/>	Fcer1g	Fc fragment of IgE receptor Ig	positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:10572	high affinity immunoglobulin epsilon receptor gamma-subunit pthr16803	gene	
<input type="checkbox"/>	Fcer1g	Fc fragment of IgE receptor Ig	positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:10572	high affinity immunoglobulin epsilon receptor gamma-subunit pthr16803	gene	
<input type="checkbox"/>	Fcer1g	Fc fragment of IgE receptor Ig	positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:10572	high affinity immunoglobulin epsilon receptor gamma-subunit pthr16803	gene	
<input type="checkbox"/>	Fcgr2a	Fc fragment of IgG receptor IIa	positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:736451	immunoglobulin fc receptor pthr11481	gene	
<input type="checkbox"/>	Fcgr2a	Fc fragment of IgG receptor IIa	positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:736451	immunoglobulin fc receptor pthr11481	gene	
<input type="checkbox"/>	Fcgr2a	Fc fragment of IgG receptor	regulation of antibody-		RGD	Rattus norvegicus	IMP		immunoglobulin fc receptor pthr11481	gene	





### Number of annotations by evidence

Species filter:



## GO Enrichment Analysis ?

Powered by PANTHER

APOH  
APP  
CND2  
COL3A1  
COL5A2  
CXCL6  
FOXP4

biological process

Homo sapiens

Examples

Launch >

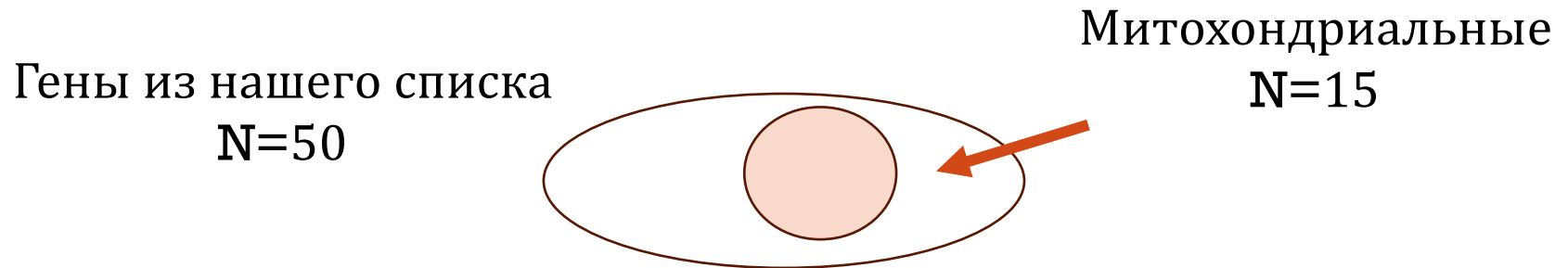
*Gene set example: genes up-regulated during formation of blood vessels/angiogenesis (source: msigdb)*

<https://geneontology.org/>

# АНАЛИЗ ОБОГАЩЕНИЯ ТЕРМИНАМИ GO

Есть список генов.

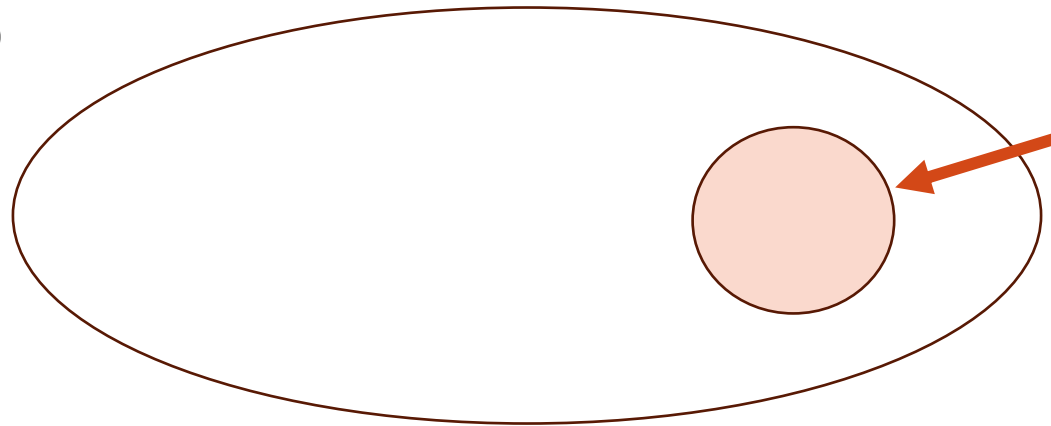
Перепредставлены ли в нем митохондриальные гены?



Это много или мало? С чем сравнить?

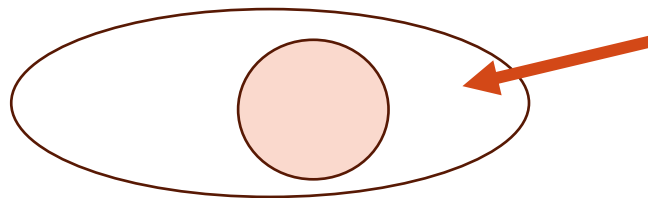
Все гены в геноме  
N=20 000

Митохондриальные  
N=50



Гены из нашего  
списка  
N=50

Митохондриальные  
N=15

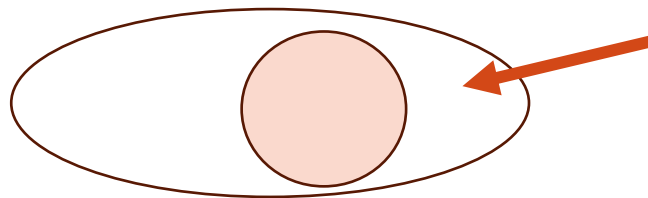


Все гены в геноме  
N=20 000

Митохондриальные  
N=50

Насколько чаще гены из нашего списка  
попадают в категорию «митохондриальные»,  
чем если бы мы их выбрали из всех генов  
случайно?

Гены из нашего  
списка  
N=50



Митохондриальные  
N=15

# АНАЛИЗ ОБОГАЩЕНИЯ ТЕРМИНАМИ GO

- Для GO термина «митохондриальные» можно все гены разделить на:
  - Гены из нашего списка из GO термина «митохондриальные»
  - Гены из нашего списка вне GO термина «митохондриальные»
  - Все гены из базы (background) из GO термина «митохондриальные»
  - Все гены из базы (background) вне GO термина «митохондриальные»



# АНАЛИЗ ОБОГАЩЕНИЯ ТЕРМИНАМИ GO

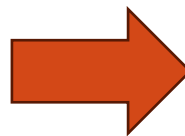
- В GO тысячи терминов!



# РАЗНЫЕ СЕРВИСЫ ИСПОЛЬЗУЮТ РАЗНЫЕ СТАТИСТИЧЕСКИЕ ТЕСТЫ

(Выбор теста зависит от задачи)

- Точный тест Фишера
- Биномиальный тест
- Критерий хи-квадрат
- и другие



P-value

**Selection Summary:**

**Analysis Type:** PANTHER Overrepresentation Test (Released 20200407)


**Annotation Version and Release Date:** GO Ontology database Released 2020-02-21

**Analyzed List:** upload\_1 (Homo sapiens)

[Change](#)

**Reference List:** Homo sapiens (all genes in database)

[Change](#)

**Annotation Data Set:**  

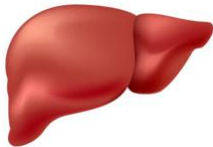
**Test Type:**  Fisher's Exact  Binomial

**Correction:**  Calculate False Discovery Rate  Use the Bonferroni correction for multiple testing   No correction

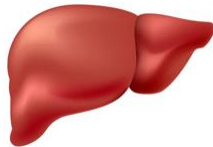
[Launch analysis](#)

# СМОТРИМ ДИФФЕРЕНЦИАЛЬНУЮ ЭКСПРЕССИЮ В ОБРАЗЦАХ СПЕЦИФИЧЕСКОЙ ТКАНИ

Sample 1



Sample 2



Список генов



GO Enrichment

- Сравниваем частоты терминов GO с частотами по всем генам
- Какие термины будут обогащены?

Можно указать свой список генов, с которыми будут сравниваться частоты терминов  
(background list)

Selection Summary:

**Analysis Type:** PANTHER Overrepresentation Test (Released 20240226)

---

**Annotation Version and Release Date:** GO Ontology database DOI: 10.5281/zenodo.10516401 Released 2024-01-17

---

**Analyzed List:** upload\_1 (Homo sapiens) [Change](#)

---

**Reference List:** Homo sapiens (all genes in database) [Change](#)

---

**Annotation Data Set:** GO biological process complete [?](#)

---

**Test Type:**  Fisher's Exact  Binomial

---

**Correction:**  Calculate False Discovery Rate  Use the Bonferroni correction for multiple testing [?](#)  No correction

[Launch analysis](#)

Для чего это нужно?



**GORILLA**



*Gene Ontology enRIchment anaLysis and visuaLizAtion tool*

**Step 1: [Choose organism](#)** ▼**Step 2: [Choose running mode](#)**

- Single ranked list of genes     Two unranked lists of genes (target and background lists)

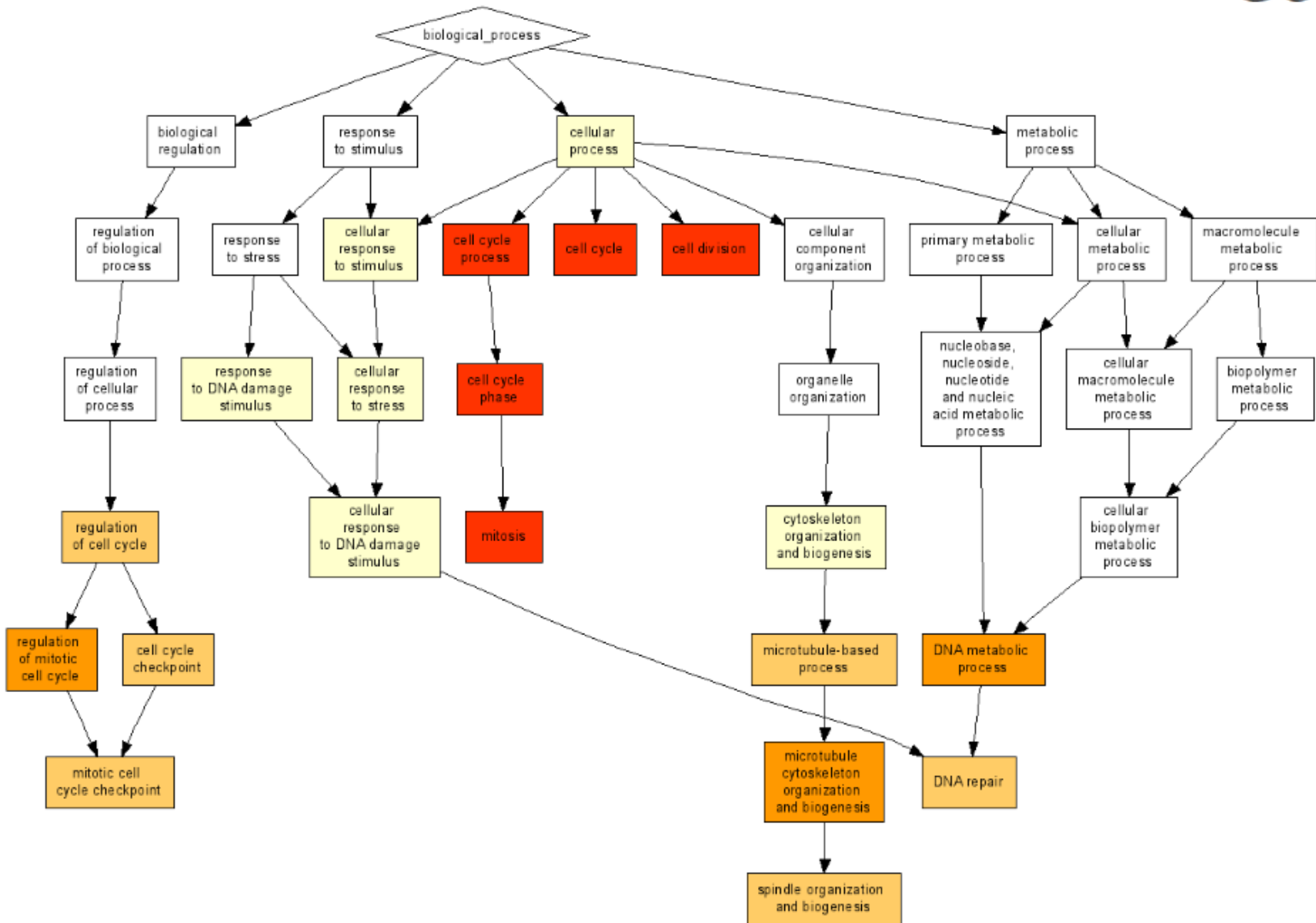
**Step 3: [Paste a ranked list of gene/protein names](#)**

Names should be separated by an <ENTER>. The preferred format is gene symbol. Other supported formats are: gene and protein RefSeq, Uniprot, Unigene and Ensembl.

Or upload a file:  Файл не выбран

**Step 4: [Choose an ontology](#)**

- Process     Function     Component     All





## 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
- ✓ Link gene-disease associations
- ✓ Highlight protein functional domains and motifs
- ✓ Redirect to related literatures
- ✓ Convert gene identifiers from one type to another.
- ✓ And more

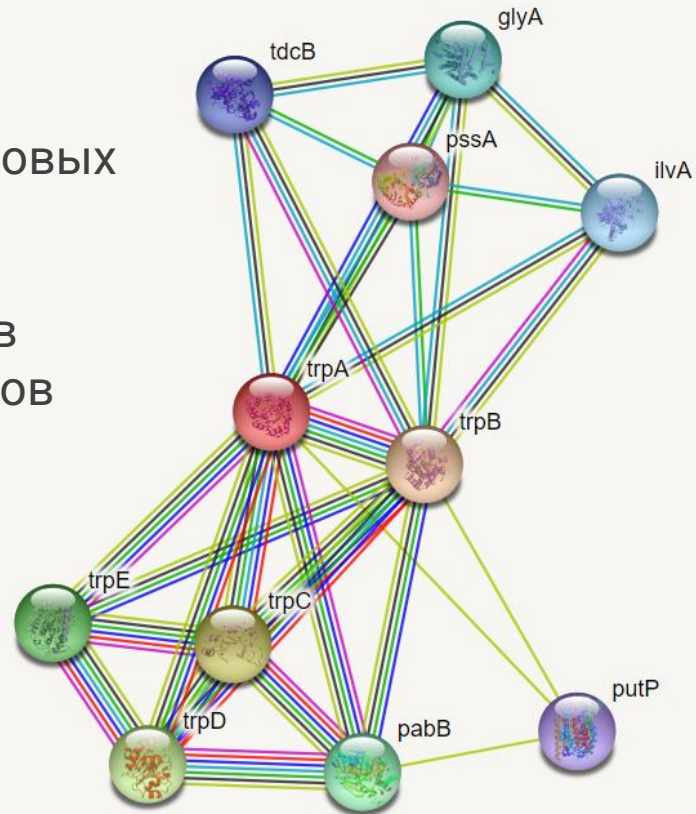
Только человек

Можно кластеризовать  
похожие термины аннотации

Можно смотреть ассоциации  
генов с болезнями



- Изучение белок-белковых взаимодействий
- 59 миллионов белков
- ~ 12 тысяч организмов



<https://string-db.org/>

### 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*

### Known Interactions

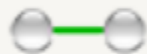


*from curated databases*



*experimentally determined*

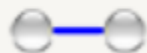
### Predicted Interactions



*gene neighborhood*



*gene fusions*



*gene co-occurrence*

### Others



*textmining*



*co-expression*



*protein homology*

Protein by name >

Protein by sequence >

**Multiple proteins** >

Multiple sequences >

Proteins with Values/Ranks <sup>New</sup> >

Organisms >

Protein families ("COGs") >

Examples >

Random entry >

## SEARCH

### Multiple Proteins by Names / Identifiers

List Of Names: (one per line; examples: [#1](#) [#2](#) [#3](#))

COQ6  
CUBN  
CYP24A1  
CYP27A1  
CYP27B1  
CYP2E1

... or, upload a file:

Browse ...

Organism:

Homo sapiens

SEARCH

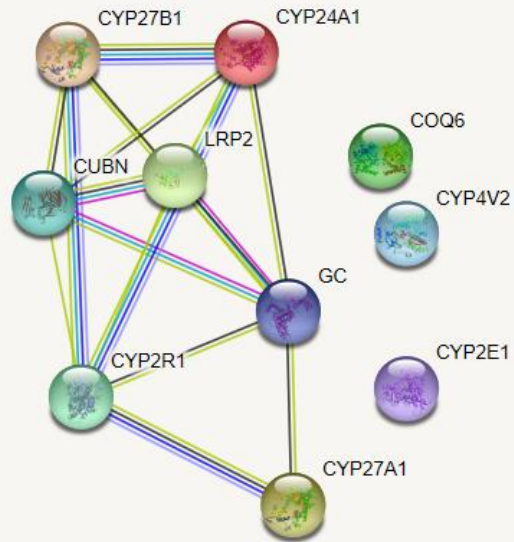
'COQ6':

- COQ6** - Ubiquinone biosynthesis monooxygenase **COQ6**, mitochondrial; FAD-dependent monooxygenase required for the C5-ring hydroxylation during ubiquinone biosynthesis. Catalyzes the hydroxylation of 3-polyprenyl-4-hydroxybenzoic acid to 3- polyprenyl-4,5-dihydroxybenzoic acid. The electrons required for the hydroxylation reaction may be funneled indirectly from NADPH via a ferredoxin/ferredoxin reductase system to **COQ6**

'CUBN':

- CUBN** - Cubilin; Cotransporter which plays a role in lipoprotein, vitamin and iron metabolism, by facilitating their uptake. Binds to ALB, MB, Kappa and lambda-light chains, TF, hemoglobin, GC, SCGB1A1, APOA1, high density lipoprotein, and the GIF-cobalamin complex. The binding of all ligands requires calcium. Serves as important transporter in several absorptive epithelia, including intestine, renal proximal tubules and embryonic yolk sac. Interaction with LRP2 mediates its trafficking throughout vesicles and facilitates the uptake of specific ligands like GC, hemoglobin, ALB, TF and SCGB1A1. [...]
- GIF - Gastric intrinsic factor; Promotes absorption of the essential vitamin cobalamin (Cbl) in the ileum. After interaction with **CUBN**, the GIF-cobalamin complex is internalized via receptor-mediated endocytosis [*a.k.a. IFMH, ENST00000533847, CCDS7977*]
- LRP2 - Low-density lipoprotein receptor-related protein 2; Multiligand endocytic receptor (By similarity). Acts together with **CUBN** to mediate endocytosis of high-density lipoproteins (By similarity). Mediates receptor-mediated uptake of polybasic drugs such as aprotinin, aminoglycosides and polymyxin B (By similarity). In the kidney, mediates the tubular uptake and clearance of leptin (By similarity). Also mediates transport of leptin across the blood-brain barrier through endocytosis at the choroid plexus epithelium (By similarity). Endocytosis of leptin in neuronal cells is required for hyp [...] [*a.k.a. HPA005980, GP330, NP\_004516.2*]
- AMN - Protein amnionless; Necessary for efficient absorption of vitamin B12. Required for normal **CUBN**- mediated protein transport in the kidney. May direct the production of trunk mesoderm during development by modulating a bone morphogenetic protein (BMP) signaling pathway in the underlying visceral endoderm (By similarity) [*a.k.a. UNQ513/PRO1028, NP\_112205.2, OTTHUMT00000415706*]

'CYP24A1':



- Viewers >
- Legend >
- Settings ▾
- Analysis >
- Exports >
- Clusters >
- + More
- Less

### Basic Settings

#### meaning of network edges:

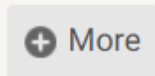
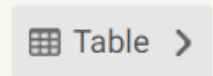
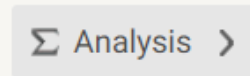
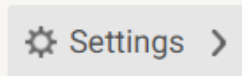
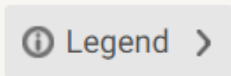
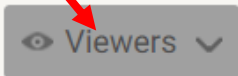
- evidence (  line color indicates the type of interaction evidence )
- confidence (  line thickness indicates the strength of data support )
- molecular action (  line shape indicates the predicted mode of action )

#### active interaction sources:

- Textmining
- Experiments
- Databases
- Co-expression
- Neighborhood
- Gene Fusion
- Co-occurrence

UPDATE





### Network

Summary view: shows current interactions. Nodes can be moved; popups provide information on nodes & edges.



### Cooccurrence

Gene families whose occurrence patterns across genomes show similarities.



### Experiments

Co-purification, co-crystallization, Yeast2Hybrid, Genetic Interactions, etc ... as imported from primary sources.



### Coexpression

Proteins whose genes are observed to be correlated in expression, across a large number of experiments.



### Databases

Known metabolic pathways, protein complexes, signal transduction pathways, etc ... from curated databases.



### Neighborhood

Groups of genes that are frequently observed in each other's genomic neighborhood.



### Textmining

currently showing

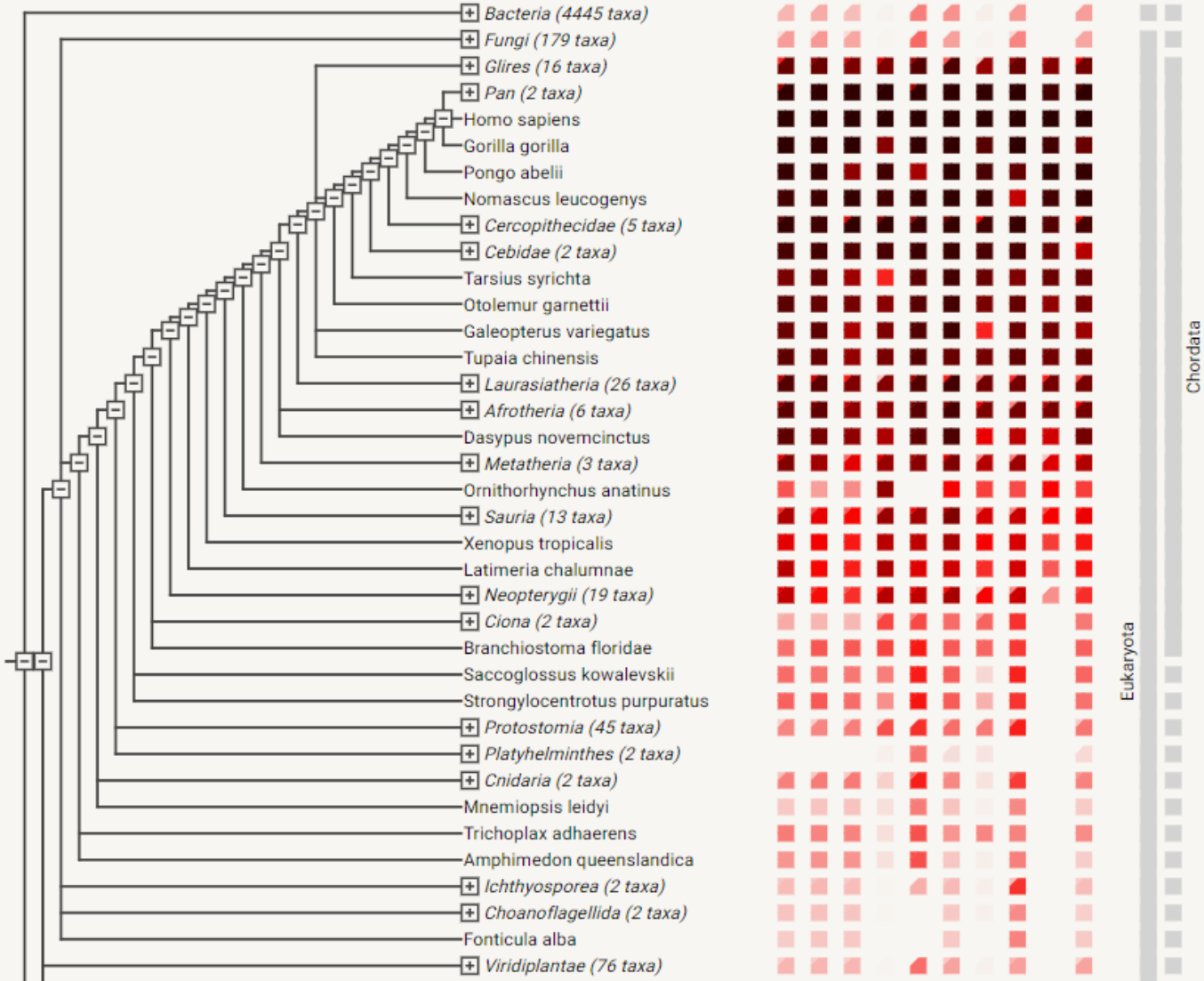
Automated, unsupervised textmining - searching for proteins that are frequently mentioned together.



### Fusion

Genes that are sometimes fused into single open reading frames.

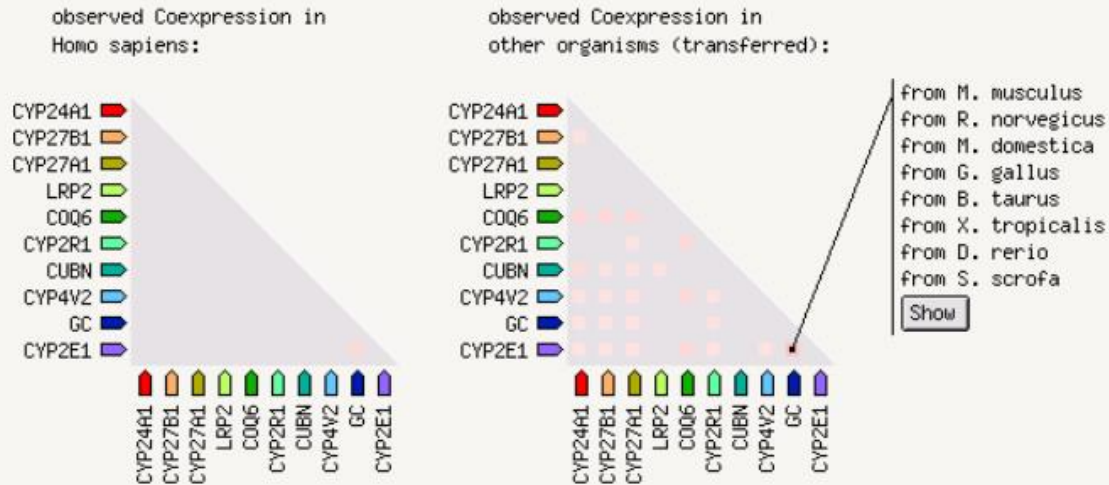
CYP24A1  
 CYP27B1  
 CYP27A1  
 LRP2  
 COQ6  
 CYP2R1  
 CUBN  
 CYP4V2  
 GC  
 CYP2E1



Chordata


Eukaryota

# GENE COEXPRESSION




Coexpression scores based on RNA expression patterns and protein co-regulation provided by [ProteomeHD](#)  
[click on the heatmap elements for details]


Viewers ▾
Legend >
Settings >
Σ Analysis >
Table >
+ More
- Less




**Network**  
Summary view: shows current interactions. Nodes can be moved; popups provide information on nodes & edges.



**Cooccurrence**  
Gene families whose occurrence patterns across genomes show similarities.



**Experiments**  
Co-purification, co-crystallization, Yeast2Hybrid, Genetic Interactions, etc ... as imported from primary sources



**Coexpression** currently showing  
Proteins whose genes are observed to be correlated in expression, across a large number of experiments.



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

Было 2 года назад

# THE HUMAN PROTEIN ATLAS

≡ MENU HELP NEWS

SEARCH<sup>1</sup>

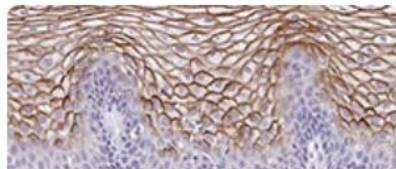
Search

Fields »

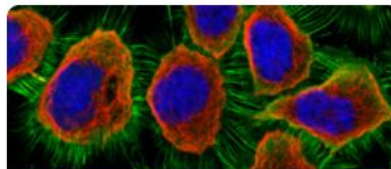
e.g. RBM3, insulin, CD36



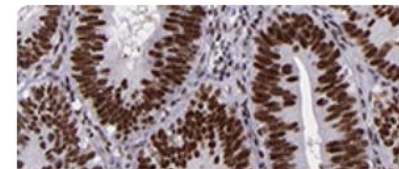
SARS-CoV-2 relevant proteins



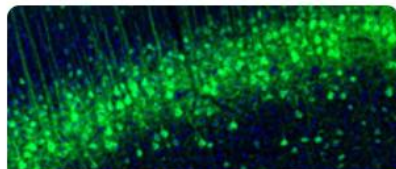
TISSUE ATLAS



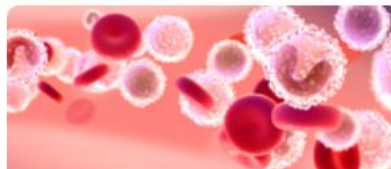
CELL ATLAS



PATHOLOGY ATLAS



BRAIN ATLAS



BLOOD ATLAS



METABOLIC ATLAS

# THE HUMAN PROTEIN ATLAS



SECTIONS ABOUT NEWS LEARN DATA HELP

А теперь...

The open access resource for human proteins

Search for specific genes/proteins or explore the 12 different sections

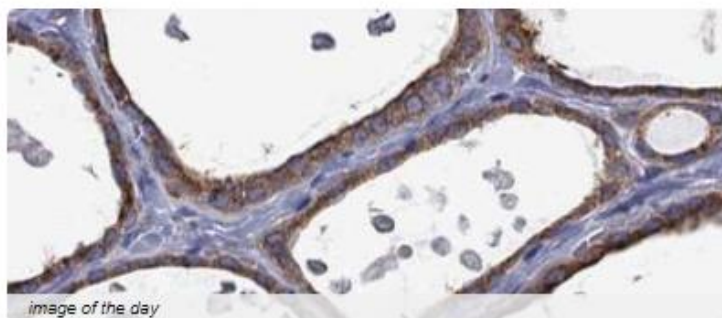
 [Fields »](#)  
[Search help](#)

image of the day



## Exploring CLOCK: Our Inner Timekeeper

Our cells contain tiny molecular clocks that control a multitude of physiological and behavioral processes according to the circadian rhythm.



Wed, 27 Mar 2024

## Different expression patterns in invasive and noninvasive PitNETs

In a study published in the Journal of the Endocrine Society, transcriptomics analysis and radiological evaluation was used to explore the differences between invasive and noninvasive pituitary neuroendocrine tumors (PitNETs) with respect to parasellar growth, bone invasiveness and vascularization.



Thu, 4 Apr 2024

[all news articles](#)

**TISSUE**

Protein and RNA profiles in tissues based on antibodies and transcriptomics

**BRAIN**

Protein and RNA profiles in brain based on microdissected regions

**SINGLE CELL TYPE**

Single cell RNA profiles in tissues based on single cell RNA analysis

**TISSUE CELL TYPE**

Cell type profiles in tissues based on deconvolution of bulk transcriptomics

**PATHOLOGY**

Protein and RNA profiles in human cancers based on antibodies and transcriptomics

**DISEASE**

Protein levels in blood in patients with diseases

**IMMUNE CELL**

RNA profiles in human immune cells

**BLOOD PROTEIN**

Blood protein levels in healthy individuals and the human secretome

**SUBCELLULAR**

Spatial, subcellular protein profiles in human cells based on antibodies

**CELL LINE**

RNA profiles in human cell lines with best models for human cancers

**STRUCTURE**

3D-structures of human proteins with antibody-binding sites and genetic variants

**INTERACTION**

Human protein-protein interactions and metabolic enzyme profiles

ACE2

Search

Fields »

Search result (19 genes): ACE2 | BACE2 | RCE1 ...

ACE2



SUMMARY



TISSUE



RNA

BRAIN



RNA

SINGLE CELL



RNA

TISSUE CELL



PATHOLOGY



PEA

DISEASE



RNA

IMMUNE



IA

BLOOD



RNA

SUBCELL



RNA

CELL LINE



STRUCTURE



INTERACTION

Ваш ID может быть специфичен для какой-либо ткани, может присутствовать везде или нигде не быть детектирован

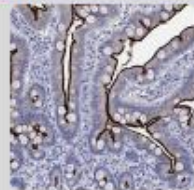
PROTEIN SUMMARY

SECTION OVERVIEW

GENE INFORMATION

RNA DATA

ANTIBODY DATA

TISSUE<sup>1</sup>Tissue expression cluster<sup>1</sup>

Kidney &amp; Intestine - Metabolism (mainly)

Tissue specificity<sup>1</sup>

Tissue enhanced (gallbladder, intestine, kidney)

Membranous expression in proximal renal tubules, intestinal tract, seminal vesicle, epididymis, exocrine pancreas and gallbladder. Expressed in Sertoli and Leydig cells, and trophoblasts. Membranous expression in ciliated cells in nasal mucosa, bronchus, and fallopian tube. Expressed in endothelial cells and pericytes in many tissues.

Protein expression<sup>1</sup>BRAIN<sup>1</sup>Human regional specificity<sup>1</sup>

Low region specificity

Pig regional specificity<sup>1</sup>

Not detected

Mouse regional specificity<sup>1</sup>

Low region specificity

SINGLE CELL TYPE<sup>1</sup>Single cell type expression cluster<sup>1</sup>

Proximal enterocytes - Digestion (mainly)

Cell type specificity<sup>1</sup>

Cell type enriched (Proximal enterocytes)

TISSUE CELL TYPE<sup>1</sup>

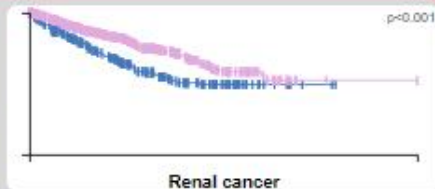
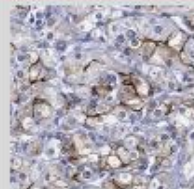
Proximal tubular cells



Hepatocytes

Classification<sup>1</sup>

Cell type enriched in 2 tissues

PATHOLOGY<sup>1</sup>

Renal cancer

Prognostic summary

Prognostic marker in renal cancer (favorable) and liver cancer (favorable)

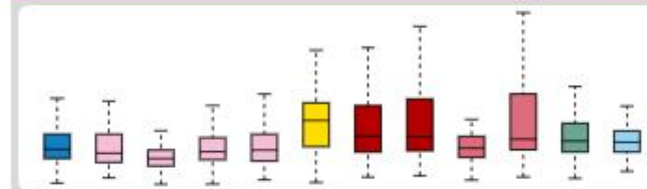
Cancer specificity<sup>1</sup>

Group enriched (colorectal cancer, renal cancer)

Cancer distribution<sup>1</sup>

Detected in many

## DISEASE

Upregulated in disease<sup>1</sup>

Glioma, Diffuse large B-cell lymphoma, Acute myeloid leukemia, Lung cancer

Disease prediction model<sup>1</sup>

No



# ACE2



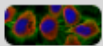
## SUBCELLULAR

### RNA EXPRESSION

Dictionary



Subcellular proteome

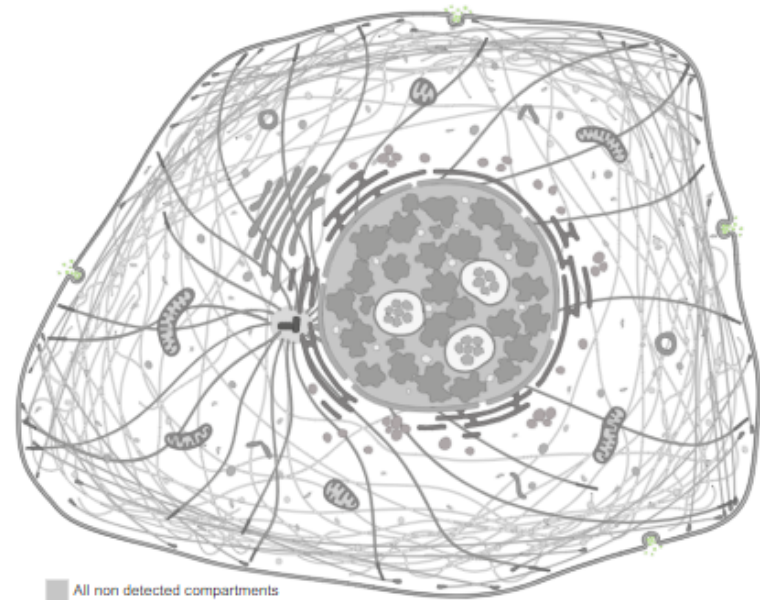


### GENERAL INFORMATION<sup>1</sup>

Gene name <sup>1</sup>	ACE2
Gene description <sup>1</sup>	Angiotensin converting enzyme 2
Protein class <sup>1</sup>	Enzymes Metabolic proteins Transporters
Predicted location <sup>1</sup>	Secreted, Membrane
Number of transcripts <sup>1</sup>	9

### HUMAN PROTEIN ATLAS INFORMATION<sup>1</sup>

Main location <sup>1</sup>	Not available (Failed to get Reactome Data)
Extracellular location <sup>1</sup>	Protein predicted to be secreted
Reliability score <sup>1</sup>	Pending cell analysis.



■ All non detected compartments  
■ ACE2 is predicted to be secreted

## TISSUE

PRIMARY DATA

TISSUES

ANTIBODIES  
AND  
VALIDATION

Dictionary

Tissue proteome

RNA AND PROTEIN EXPRESSION SUMMARY<sup>1</sup>

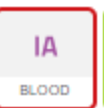
Expression Detection All organs

RNA expression (nTPM)<sup>1</sup>Protein expression (score)<sup>1</sup>

Organ/Tissue	RNA expression (nTPM) <sup>1</sup>	Protein expression (score) <sup>1</sup>	Image
Brain	High	High	Cerebral cortex
Eye	High	High	Nasopharynx
Endocrine tissues	High	High	Nasopharynx
Respiratory system	High	High	Colon
Proximal digestive tract	High	High	Colon
Gastrointestinal tract	High	High	Liver
Liver & Gallbladder	High	High	Liver
Pancreas	High	High	Kidney
Kidney & Urinary bladder	High	High	Kidney
Male tissues	High	High	Testis
Female tissues	High	High	Testis
Muscle tissues	High	High	Lymph node
Connective & Soft tissue	High	High	Lymph node
Skin	High	High	Lymph node
Bone marrow & Lymphoid tissues	High	High	Lymph node

Информация об  
экспрессии  
РНК И(!!!) белка

# ACE2



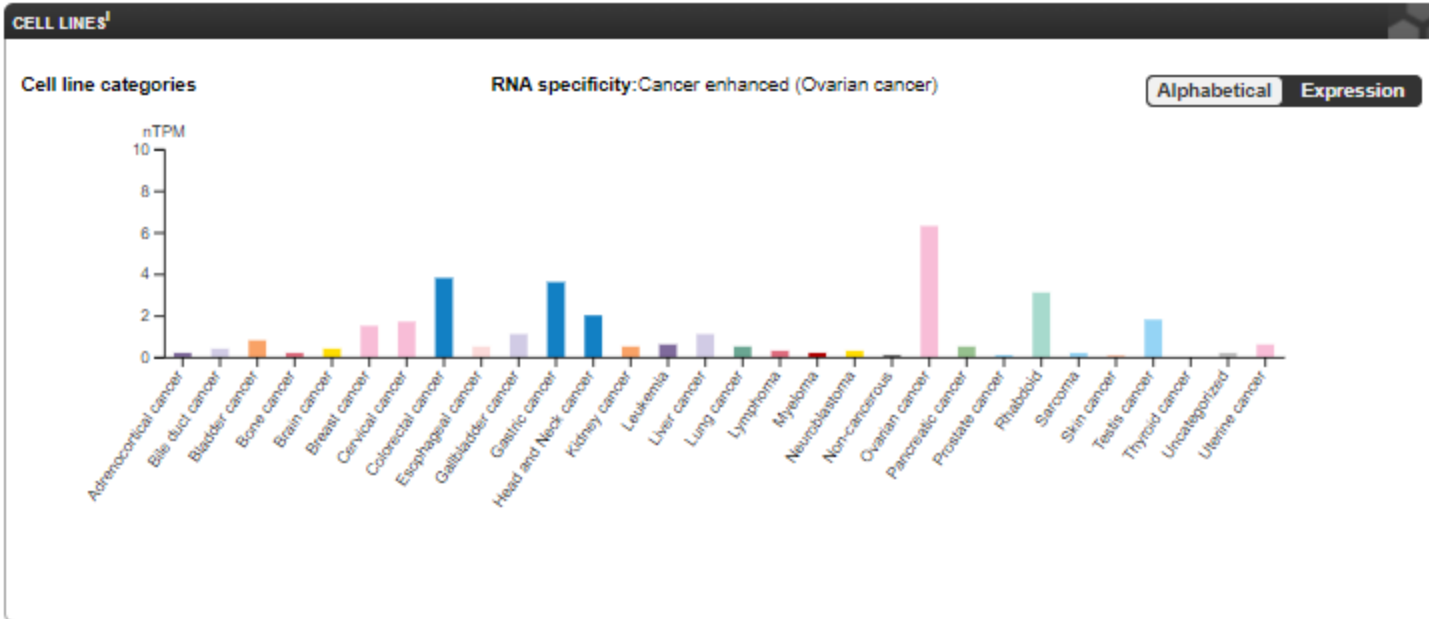
## CELL LINE

Human cell lines



GENERAL INFORMATION <sup>1</sup>	
Gene name <sup>1</sup>	ACE2
Gene description <sup>1</sup>	Angiotensin converting enzyme 2
Protein class <sup>1</sup>	Enzymes Metabolic proteins Transporters
Predicted location <sup>1</sup>	Secreted, Membrane
Number of transcripts <sup>1</sup>	9

HUMAN PROTEIN ATLAS INFORMATION <sup>1</sup>	
Cell line expression cluster <sup>1</sup>	Non-specific - Mitochondria (mainly)
Cell line specificity <sup>1</sup>	Cancer enhanced (Ovarian cancer)
Tau specificity score <sup>1</sup>	0.72
Cell line distribution <sup>1</sup>	Detected in many
Protein evidence <sup>1</sup>	Evidence at protein level
Extracellular location <sup>1</sup>	Protein predicted to be secreted





## PROTEIN STRUCTURE

ANTIBODIES AND VALIDATION



Protein structures



### GENERAL INFORMATION<sup>1</sup>

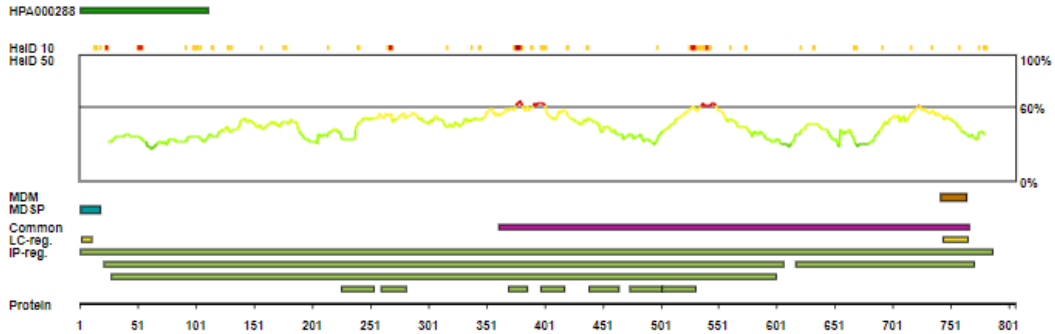
Gene name <sup>1</sup>	ACE2
Synonyms	ACEH
Gene description <sup>1</sup>	Angiotensin converting enzyme 2
Protein class <sup>1</sup>	Enzymes Metabolic proteins Transporters
Predicted location <sup>1</sup>	Secreted, Membrane
Protein evidence	Evidence at protein level ( <a href="#">all genes</a> )

### GENE INFORMATION<sup>1</sup>

Chromosome	X
Cytoband	p22.2
Chromosome location (bp)	15494566 - 15607236
Number of transcripts <sup>1</sup>	9
Ensembl	ENSG00000130234 (version 109)
Entrez gene	59272
HGNC	HGNC:13557
UniProt	Q9BYF1 (UniProt - Evidence at protein level)
neXtProt	NX_Q9BYF1
GeneCards	ACE2

### PROTEIN BROWSER<sup>1</sup>

ACE2-201 ACE2-202 ACE2-205 ACE2-206 ACE2-207 ACE2-208 ACE2-209 ACE2-210 ACE2-211



#### Description:

Structure prediction of Q9BYF1 from AlphaFold project, version 2

#### Color scheme:

Confidence  Residue index  Your selection

#### Variants:

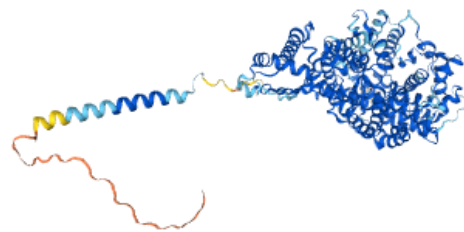
Off  Clinical (#=0)  Population (#=256)

#### Autorotate:

Off  On

#### Confidence for predicted structure:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)





## INTERACTION

Protein Interactions



### GENERAL INFORMATION<sup>1</sup>

Gene name <sup>1</sup>	ACE2
Gene description <sup>1</sup>	Angiotensin converting enzyme 2
Protein class <sup>1</sup>	Enzymes Metabolic proteins Transporters
Predicted location <sup>1</sup>	Secreted, Membrane
Number of transcripts <sup>1</sup>	9

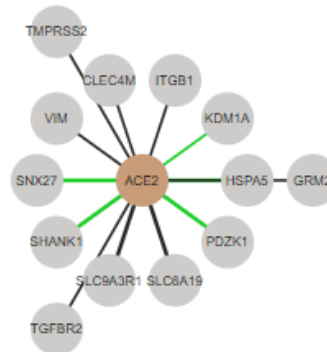
### HUMAN PROTEIN ATLAS INFORMATION<sup>1</sup>

Tissue expression cluster (RNA) <sup>1</sup>	Kidney & Intestine - Metabolism (mainly)
Tissue specificity (RNA) <sup>1</sup>	Tissue enhanced (gallbladder, intestine, kidney)
Tau specificity score (RNA) <sup>1</sup>	0.77
Tissue distribution (RNA) <sup>1</sup>	Detected in many
Protein evidence <sup>1</sup>	Evidence at protein level
Subcellular location <sup>1</sup>	Not available
Secretome annotation <sup>1</sup>	Secreted to blood

### INTERACTION SUMMARY<sup>1</sup>

Filter menu »

Highlight:  Off  Subcell location  Predicted location  Tissue specificity  Protein class



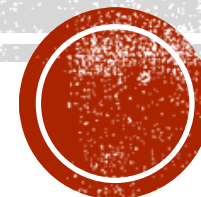
● Matching filter  
— Direct interaction  
— Physical association

Number of interactions: 13

[View all protein interaction data in IntAct](#)

Interaction	Interaction type	Confidence	MI score	# Interactions
HSPA5	Direct interaction	High	0.72	22
PDZK1	Direct interaction	High	0.71	6
SLC6A19	Physical association	High	0.64	1
SLC9A3R1	Physical association	High	0.63	16
SHANK1	Direct interaction	High	0.61	2
SNX27	Direct interaction	High	0.61	2
VIM	Physical association	Medium	0.6	88

# Метаболические пути





KEGG ▼

Search

Help

[→ Japanese](#)

#### KEGG Home

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

#### KEGG Database

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

#### KEGG Objects

[Pathway maps](#)  
[Brite hierarchies](#)  
[KEGG DB links](#)

#### KEGG Software

[KEGG API](#)  
[KGML](#)

#### KEGG FTP

[Subscription](#)

[GenomeNet](#)

[DBGET/LinkDB](#)

[Feedback](#)

[Copyright request](#)

[Kanehisa Labs](#)

## KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. See [Release notes](#) (April 1, 2020) for new and updated features.

### ● Main entry point to the KEGG web service

**KEGG2** [KEGG Table of Contents](#) [[Update notes](#) | [Release history](#)]

### ● Data-oriented entry points

**KEGG PATHWAY** [KEGG pathway maps](#)  
**KEGG BRITE** [BRITE hierarchies and tables](#)  
**KEGG MODULE** [KEGG modules](#)  
**KEGG ORTHOLOGY** [KO functional orthologs](#) [[Annotation](#)]  
**KEGG GENOME** [Genomes](#) [[Pathogen](#) | [Virus](#) | [Plant](#)]  
**KEGG GENES** [Genes and proteins](#) [[SeqData](#)]  
**KEGG COMPOUND** [Small molecules](#)  
**KEGG GLYCAN** [Glycans](#)  
**KEGG REACTION** [Biochemical reactions](#) [[RModule](#)]  
**KEGG ENZYME** [Enzyme nomenclature](#)  
**KEGG NETWORK** [Disease-related network elements](#)  
**KEGG DISEASE** [Human diseases](#) [[Cancer](#)]  
**KEGG DRUG** [Drugs](#) [[New drug approvals](#)]

### Classification

[Pathway](#)  
[Brite](#)  
[Brite table](#)  
[Module](#)  
[KO \(Function\)](#)  
[Organism](#)  
[Compound](#)  
[Network](#)  
[Disease \(ICD\)](#)  
[Drug \(ATC\)](#)  
[Drug \(Target\)](#)

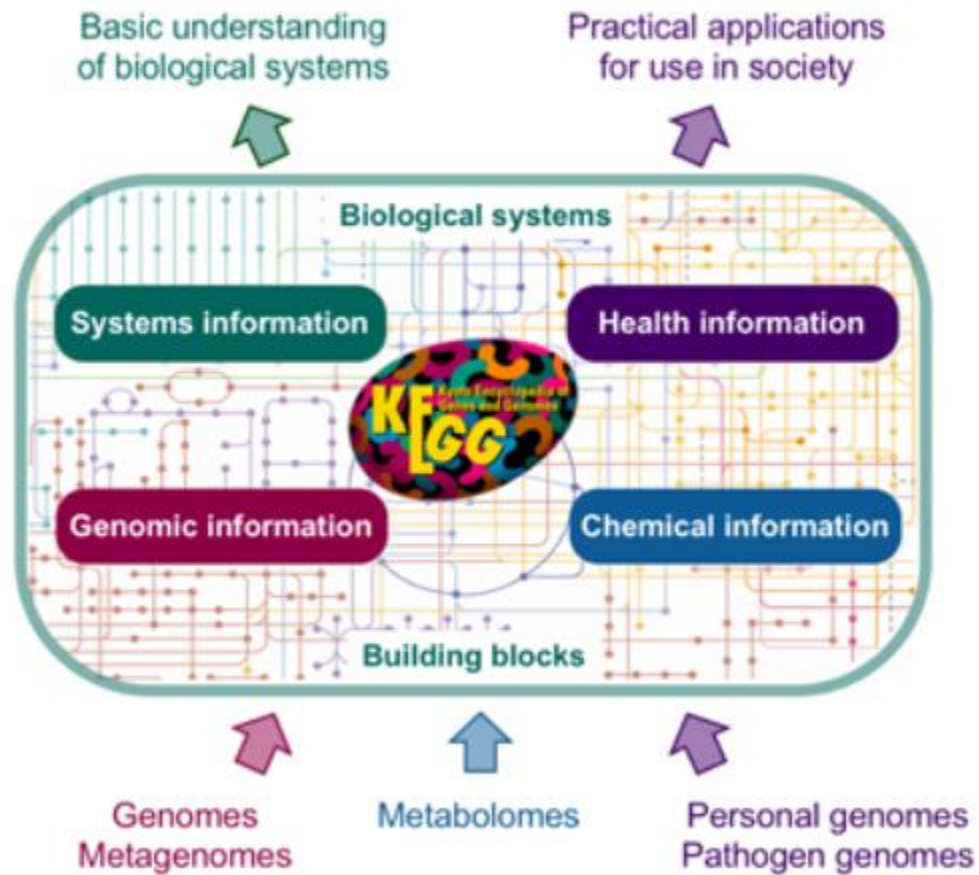
**KEGG MEDICUS** [Health information resource](#) [[Drug labels search](#)]

### ● Organism-specific entry points

**KEGG Organisms** Enter org code(s)   [hsa](#) [hsa eco](#)

### ● Analysis tools

**KEGG Mapper** [KEGG PATHWAY/BRITE/MODULE mapping tools](#)  
**BlastKOALA** [BLAST-based KO annotation and KEGG mapping](#)  
**GhostKOALA** [GHOSTX-based KO annotation and KEGG mapping](#)  
**KofamKOALA** [HMM profile-based KO annotation and KEGG mapping](#)  
**BLAST/FASTA** [Sequence similarity search](#)  
**SIMCOMP** [Chemical structure similarity search](#)





### KEGG Database as of 2024/4/3

#### Systems information

KEGG PATHWAY	Pathway maps, reference (total)	570 (1,152,602)
KEGG BRITE	Functional hierarchies, reference (total)	195 (384,000)
KEGG MODULE	KEGG modules	481
	Reaction modules	48

#### Genomic information

KEGG ORTHOLOGY	KEGG Orthology (KO) groups	26,696
KEGG GENES	Genes in KEGG organisms	52,674,035
	Addendum proteins	4,159
	Viral genes	688,823
	Viral mature peptides	377
KEGG GENOME	KEGG organisms	9,822
	(1024 eukaryotes, 8366 bacteria, 432 archaea)	
	KEGG selected viruses (T4 category)	359
	KEGG viruses (Vtax category)	14,384

#### Chemical information

KEGG COMPOUND	Metabolites and other chemical substances	19,296
KEGG GLYCAN	Glycans	11,222
KEGG REACTION	Biochemical reactions	12,016
	Reaction class	3,194
KEGG ENZYME	Enzyme nomenclature	8,109

#### Health information

KEGG NETWORK	Disease-related network elements	1,396
	Network variation maps	140
KEGG VARIANT	Human gene variants	1,328
KEGG DISEASE	Human diseases	2,705
KEGG DRUG	Drugs	12,369
	Drug groups	2,462

#### Drug labels

KEGG MEDICUS	Japanese prescription drug labels from JAPIC	13,357
	Japanese OTC drug labels from JAPIC	10,545
KEGG MEDICUS	FDA prescription drug labels linked to DailyMed	33,365



Category	Database	Content	Color
Systems information	<b>KEGG PATHWAY</b>	KEGG pathway maps	KEGG
	KEGG BRITE	BRITE hierarchies and tables	
	KEGG MODULE	KEGG modules	
Genomic information	KEGG ORTHOLOGY (KO)	Functional orthologs	KEGG
	KEGG GENOME	KEGG organisms (complete genomes)	
	KEGG GENES	Genes and proteins	KEGG
	KEGG SSDB	GENES sequence similarity	
Chemical information	KEGG COMPOUND	Small molecules	KEGG
	KEGG GLYCAN	Glycans	
	KEGG REACTION	Biochemical reactions	
	KEGG RCLASS	Reaction class	
	KEGG ENZYME	Enzyme nomenclature	
Health information	KEGG NETWORK	Disease-related network elements	KEGG
	KEGG VARIANT	Human gene variants	
	KEGG DISEASE	Human diseases	
	KEGG DRUG	Drugs	
	KEGG DGROUP	Drug groups	
	KEGG ENVIRON	Health-related substances	

Chemical information category is collectively called **KEGG LIGAND**

Health information category integrated with drug labels is called **KEGG MEDICUS**



## Pathway Maps

**KEGG PATHWAY** is a collection of manually drawn [pathway maps](#) representing our knowledge on the molecular interaction, reaction and relation networks for:

### 1. Metabolism

Global/overview   Carbohydrate   Energy   Lipid   Nucleotide   Amino acid   Other amino   Glycan  
Cofactor/vitamin   Terpenoid/PK   Other secondary metabolite   Xenobiotics   Chemical structure

### 2. Genetic Information Processing

### 3. Environmental Information Processing

### 4. Cellular Processes

### 5. Organismal Systems

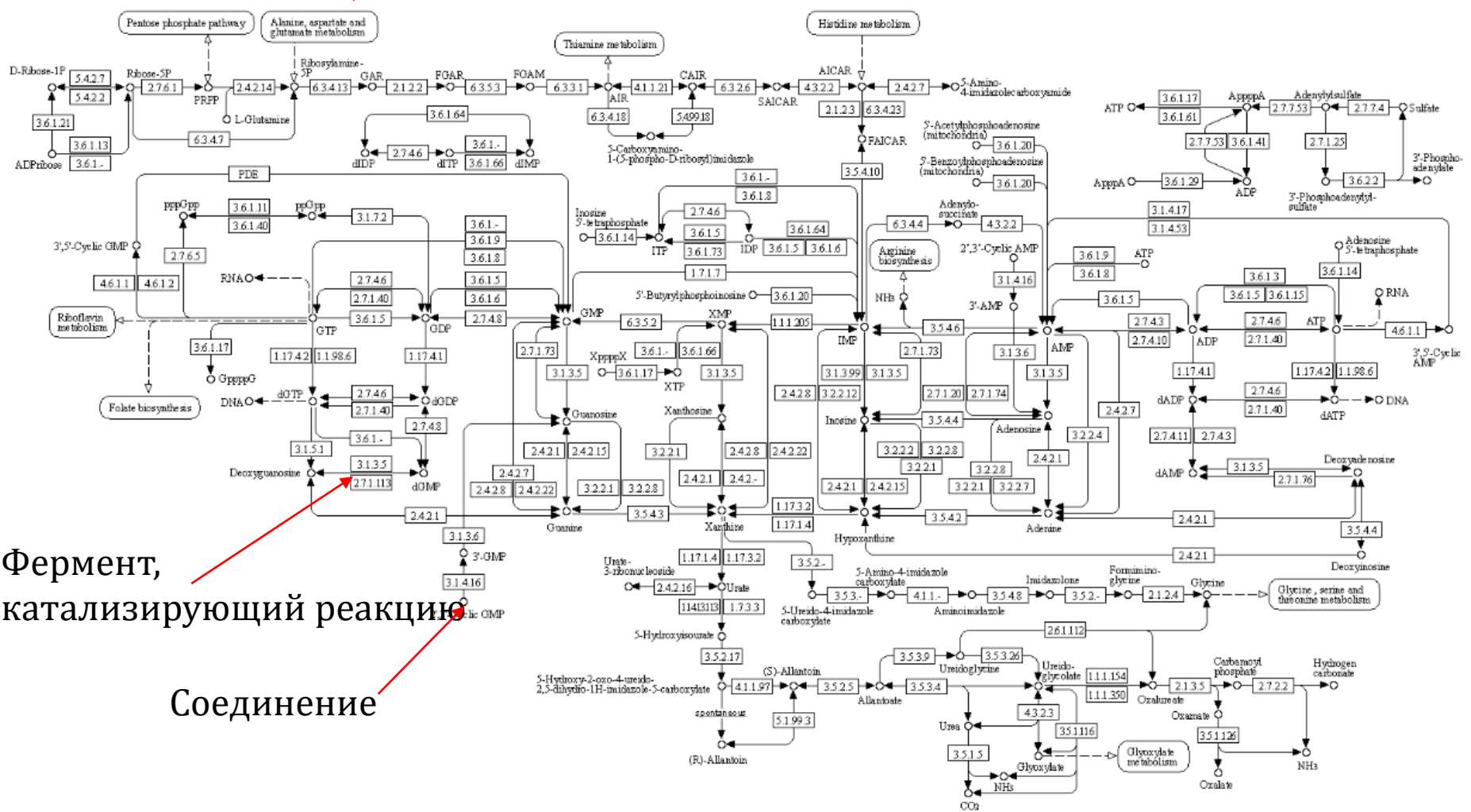
### 6. Human Diseases

### 7. Drug Development



# Другой метаболический путь

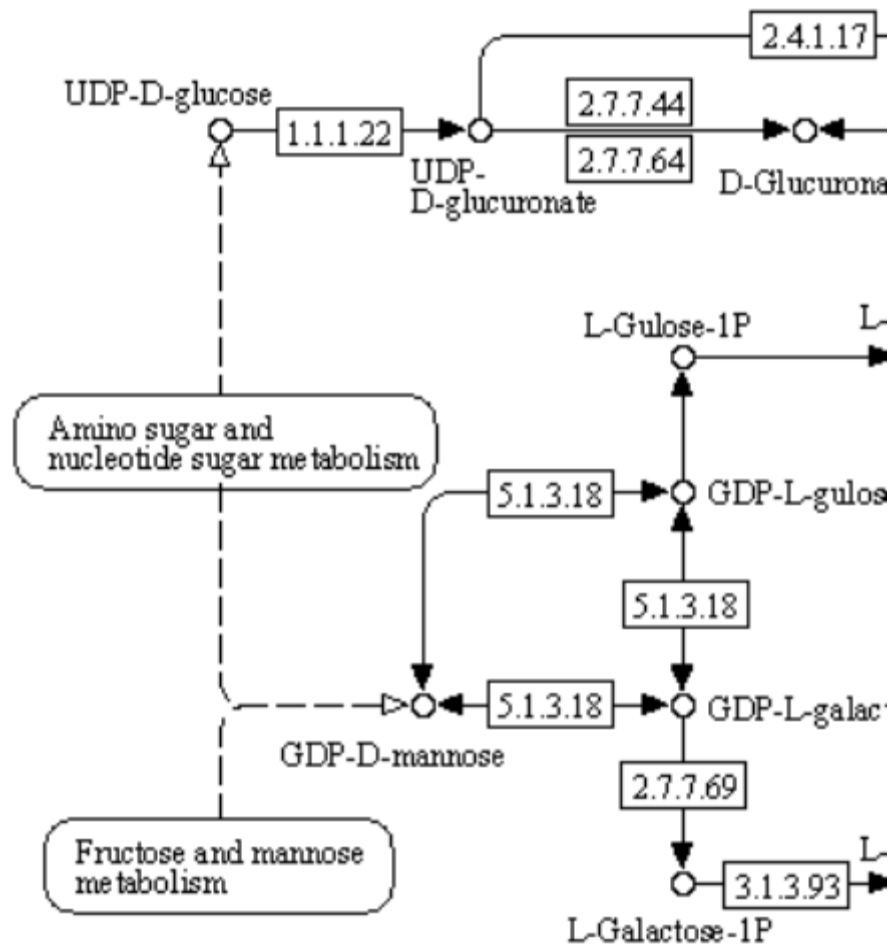
## PURINE METABOLISM



Фермент,  
катализирующий реакцию

Соединение







<b>Entry</b>	EC 2.7.7.4 Enzyme
<b>Name</b>	sulfate adenylyltransferase; ATP-sulfurylase; adenosine-5'-triphosphate sulfurylase; adenosinetriphosphate sulfurylase; adenylylsulfate pyrophosphorylase; ATP sulfurylase; ATP-sulfurylase; sulfurylase
<b>Class</b>	Transferases; Transferring phosphorus-containing groups; Nucleotidyltransferases <a href="#">BRITE hierarchy</a>
<b>Sysname</b>	ATP:sulfate adenylyltransferase
<b>Reaction(IUBMB)</b>	ATP + sulfate = diphosphate + adenylyl sulfate [RN:R00529]
<b>Reaction(KEGG)</b>	R00529; (other) R04929 <a href="#">Reaction</a>
<b>Substrate</b>	ATP [CPD:C00002]; sulfate [CPD:C00059]
<b>Product</b>	diphosphate [CPD:C00013]; adenylyl sulfate [CPD:C00224]
<b>Comment</b>	The human phosphoadenosine-phosphosulfate synthase (PAPS) system is a bifunctional enzyme (fusion product of two catalytic activities). In a first step, sulfate adenylyltransferase catalyses the formation of adenosine 5'-phosphosulfate (APS) from ATP and inorganic sulfate. The second step is catalysed by the adenylylsulfate kinase portion of 3'-phosphoadenosine 5'-phosphosulfate (PAPS) synthase, which involves the formation of PAPS from enzyme-bound APS and ATP. In contrast, in bacteria, yeast, fungi and plants, the formation of PAPS is carried out by two individual polypeptides, sulfate adenylyltransferase (EC 2.7.7.4) and adenylyl-sulfate kinase (EC 2.7.1.25).

**All links**

- Pathway (16)
  - KEGG PATHWAY (14)
  - KEGG MODULE (2)
- Chemical substance (6)
  - KEGG COMPOUND (6)
- Chemical reaction (4)
  - KEGG REACTION (2)
  - KEGG RCLASS (2)
- Gene (106612)
  - KEGG ORTHOLOGY (5)
  - KEGG GENES (8508)
  - KEGG MGENES (61373)
  - RefGene (36726)
- Protein sequence (77269)
  - UniProt (40770)
  - SWISS-PROT (457)
  - RefSeq(pep) (35969)
  - PDBSTR (54)
  - PMD (19)
- DNA sequence (141864)
  - RefSeq(nuc) (127520)
  - GenBank (7103)
  - EMBL (7241)
- 3D Structure (24)
  - PDB (24)
- Protein domain (9)
  - InterPro (8)
  - Pfam (1)
- All databases (325804)

[Download RDF](#)

# КОД ФЕРМЕНТА

- **Класс** – код первого уровня: 1-7. Номер одного из семи главных классов ферментов
- **Подкласс** – код второго уровня. Номер характеризует основные виды субстратов, участвующих в данном виде химических реакций.
- **Под-подкласс** – код третьего уровня. Номер определяет более частные подгруппы, отличающиеся природой химических соединений доноров или акцепторов, участвующих в данной подгруппе реакций.
- **Код четвертого уровня** – все ферменты, относящиеся к данному под-подклассу получают свои порядковые номера.
  
- **Пример:** глюкоксидаза – 1.1.3.4
  
- **КФ 1** — Оксидоредуктазы
- **КФ 1.1** — Алкогольоксидоредуктазы
- **КФ 1.1.3** — Оксидоредуктазы, окисляющие группу СН-ОН и восстанавливающие кислород.
- **КФ 1.1.3.4** — Оксидоредуктазы, окисляющие глюкозу в присутствии кислорода. Всего известно 3 глюкоксидазы из разных организмов.

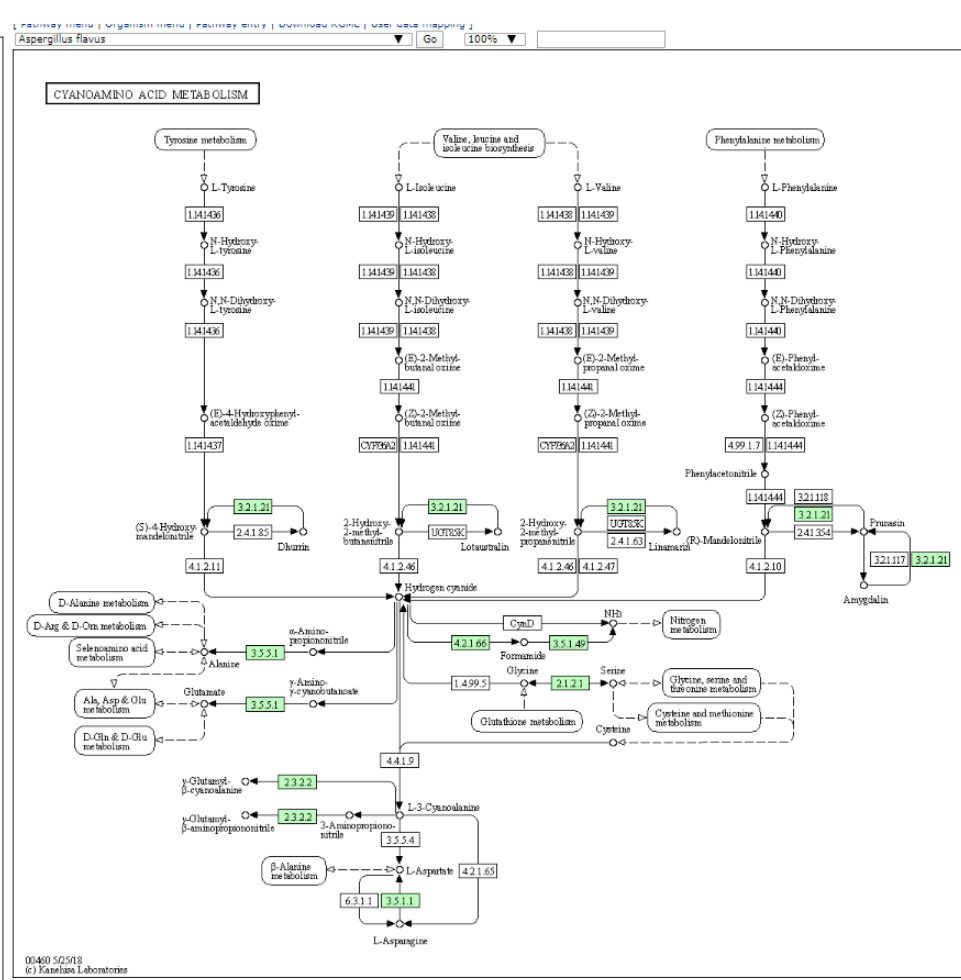
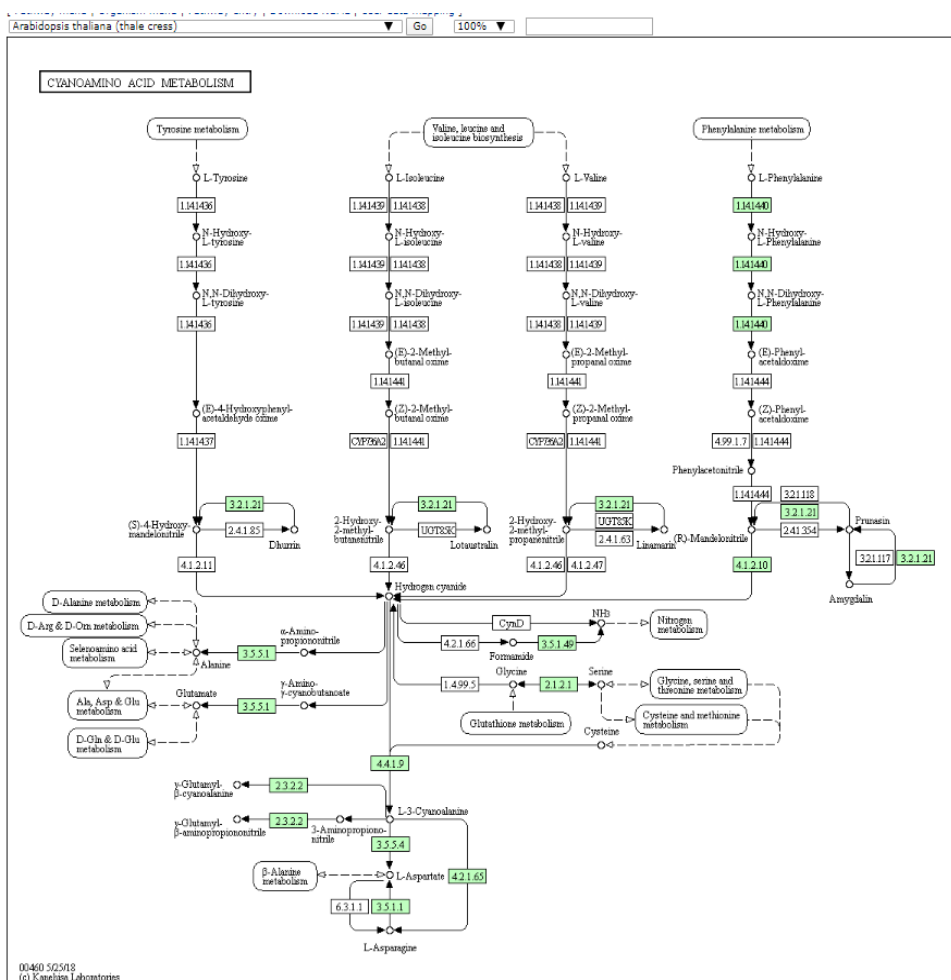
# ОСНОВНЫЕ КЛАССЫ ФЕРМЕНТОВ

## ■ Коды первого уровня


Класс	Катализируемая реакция	Тип реакции	Важнейшие подклассы
КФ 1 <i>Оксидоредуктазы</i>	Окислительно-восстановительные реакции. Перенос атомов Н и О или <b>электронов</b> от одного субстрата на другой	$AH + B \rightarrow A + BH$ (восстановленный) $A + O \rightarrow AO$ (окисленный)	дегидрогеназа, оксидаза, пероксидаза, редуктаза, монооксидаза, диоксигеназа
КФ 2 <i>Трансферазы</i>	Перенос функциональной группы от одного субстрата на другой. Это может быть метильная, ацильная, фосфатная группа или аминогруппа.	$AB + C \rightarrow A + BC$	аминотрансфераза, фосфотрансфераза, С1-трансфераза, гликозилтрансфераза
КФ 3 <i>Гидролазы</i>	Образование двух продуктов из одного субстрата в результате гидролиза.	$AB + H_2O \rightarrow AOH + BH$	эстераза, гликозил-гидролаза, пептидаза, амидаза
КФ 4 <i>Лиазы (синтазы)</i>	Негидролитическое добавление или удаление группы к или от субстрата. Образование С-С, С-Н, С-О или С-С связи.	$RCO_2COOH \rightarrow RCOH + CO_2$	С-О-лиаза, С-S-лиаза, С-N-лиаза, С-С-лиаза
КФ 5 <i>Изомеразы</i>	Внутримолекулярная перестановка, то есть <b>изомеризация</b> молекулы субстрата.	$AB \rightarrow BA$	эпимераза, цис-транс-изомераза, внутримолекулярная оксидоредуктаза и др.
КФ 6 <i>Лигазы (синтазы)</i>	Соединение двух молекул в результате синтеза новой С-О, С-S, С-N или С-С связи, сопряженное с одновременным гидролизом АТФ.	$X + Y + ATP \rightarrow XY + ADP + Pi$	С-О-лигаза, С-S-лигаза, С-N-лигаза, С-С-лигаза
КФ 7 <i>Транслоказы</i>	Перенос ионов или молекул через мембраны или их разделение в мембранах.	$AB + C_{[side\ 1]} \rightarrow AB^* + C_{[side\ 2]}$	ЕС 7.1 - транслокация H <sup>+</sup> ; ЕС 7.2 - транслокация неорганических катионов и их хелатов; ЕС 7.3 - транслокация неорганических анионов; ЕС 7.4 - транслокация аминокислот и пептидов; ЕС 7.5 - транслокация углеводов и их производных; ЕС 7.6 - транслокация других соединений. <sup>[2]</sup>



# Сравнительная геномика



# ПОИСК ОТ ОРГАНИЗМА

 Search  for

Database: KEGG - Search term: neurospora crassa

## KEGG GENES

cmi:CMM\_2688

no KO assigned | (GenBank) unnamed protein product; putative acetyl xylan esterase (XP\_330314.1) hypothetical protein [Neurospora crassa]; AAC39371.1] acetyl xylan esterase II precursor [Penicillium purpurogenum]., pfam01083, Cutinase, Cutinase.; Conserved hypothetical protein

vg:1449901

no KO assigned | (RefSeq) MSV087, MseVgp087; Melanoplus sanguinipes entomopoxvirus; ORF MSV087 putative thioredoxin, similar to Neurospora crassa GB:D45892

vg:1449935

no KO assigned | (RefSeq) MSV144, MseVgp144; Melanoplus sanguinipes entomopoxvirus; ORF MSV144 putative ubiquitin, similar to Neurospora crassa GB:U01220

## KEGG GENOME

T01034

ncr, 367110; Neurospora crassa OR74A

## KEGG MGENES

T30455:18971

similarity to hypothetical protein CAD21257.1 - neurospora crassa

T30456:15058

similarity to hypothetical protein CAD21257.1 - neurospora crassa

T30579:1733

similarity to hypothetical protein CAD21257.1 - neurospora crassa

## KEGG ENZYME

3.1.30.1

Aspergillus nuclease S1; endonuclease S1 (Aspergillus); single-stranded-nucleate endonuclease; deoxyribonuclease S1; deoxyribonuclease S1; nuclease S1; Neurospora crassa single-strand specific endonuclease; S1 nuclease; single-strand endodeoxyribonu •••



## GENOME: *Neurospora crassa*

[Help](#)

<b>Entry</b>	T01034	Complete Genome
<b>Name</b>	ncr, 367110	
<b>Definition</b>	Neurospora crassa OR74A	
<b>Category</b>	Reference genome	
<b>Annotation</b>	yes <a href="#">Show organism</a>	
<b>Taxonomy</b>	TAX:367110	
<b>Lineage</b>	Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetidae; Sordariales; Sordariaceae; Neurospora <a href="#">Taxonomy</a>	
<b>Data source</b>	RefSeq (Assembly:GCF_000182925.2) BioProject:132	
<b>Original DB</b>	Broad	
<b>Statistics</b>	Number of protein genes:	9758
	Number of RNA genes:	416
<b>Created</b>	2007	
<b>Reference</b>	PMID:12712197	
<b>Authors</b>	Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, et al.	
<b>Title</b>	The genome sequence of the filamentous fungus <i>Neurospora crassa</i> .	
<b>Journal</b>	Nature 422:859-68 (2003) DOI:10.1038/nature01554	

### All links

Ontology (43)  
KEGG BRITE (43)  
Pathway (297)  
KEGG PATHWAY (119)  
KEGG MODULE (178)  
Genome (1)  
ASSEMBLY (1)  
Gene (15422)  
KEGG GENES (10174)  
KEGG MGENES (613)  
RefGene (4635)  
Literature (1)  
PubMed (1)  
Taxonomy (1)  
TAX (1)  
All databases (15765)

[Download RDF](#)

# REACTOME

- <https://reactome.org/>



## Pathway Browser

Visualize and interact with Reactome biological pathways



## Analysis Tools

Merges pathway identifier mapping, over-representation, and expression analysis



## ReactomeFIViz

Designed to find pathways and network patterns related to cancer and other types of diseases



## Documentation

Information to browse the database and use its principal tools for data analysis



 Analysis tools



Analyse gene list



Analyse gene expression



Species Comparison



Your data

Options

Analysis

Step 1: Select a file from your computer or paste your own data and click on the corresponding "Continue" button.

Select data file for analysis:  Файл не выбран

Paste your data to analyse or try example data sets:

```

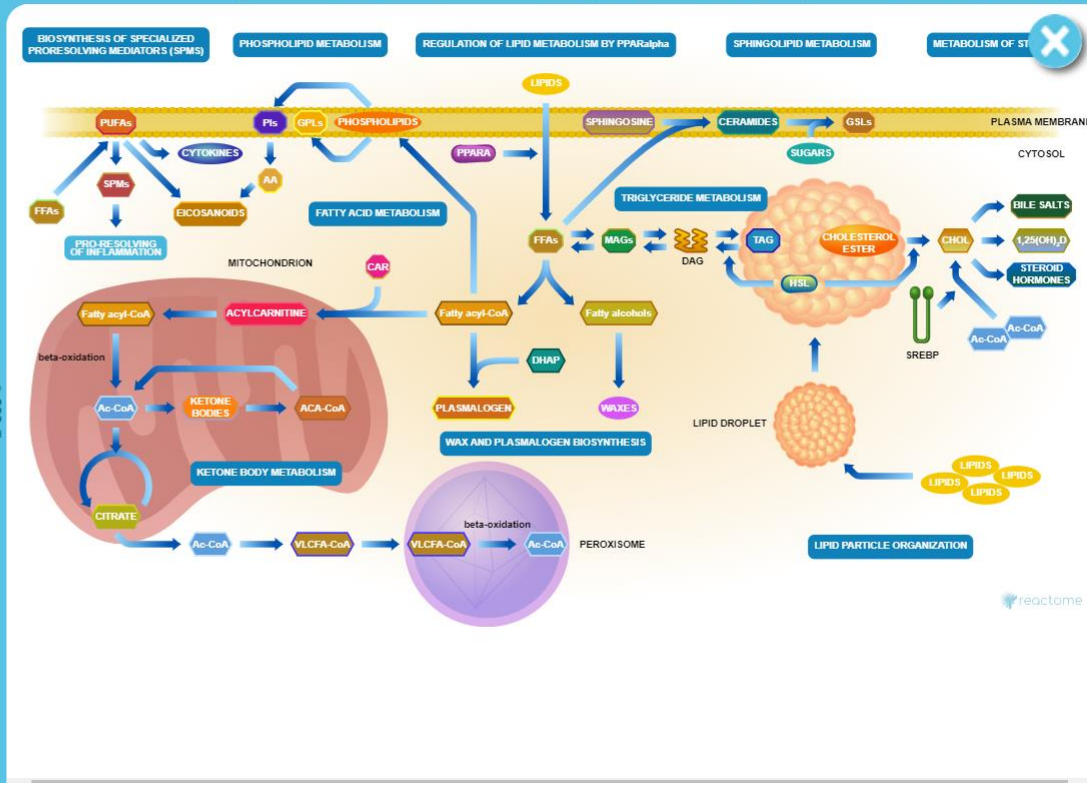
LOC100047604
D730039F16Rik
Scyl1
1700023B02Rik
1700034H14Rik
Fbxo8
Paip1
Tmem186
Atpaf1
LOC100046254
LOC100047604
Coq10a
Fn3k
Sipa111
Slc25a16
Slc25a40
Rps6ka5
Trim37
    
```

Some examples:



Event Hierarchy:

- Metabolism (77/3,634) FDR: 9.99E-1
  - Metabolism of carbohydrates (10/456) FDR: 8.38E-1
  - Inositol phosphate metabolism (1/90) FDR: 9.24E-1
  - Metabolism of lipids (21/1,437) FDR: 10E-1**
    - Integration of energy metabolism (3/145) FDR: 7.82E-1
      - Regulation of insulin secretion (3/106) FDR: 6.62E-1
      - Glucagon signaling in metabolic regulation (2/40) FDR: 6.62E-1
      - PKA-mediated phosphorylation of key metabolic factors (2/7) FDR: 6.62E-1
      - Insulin effects increased synthesis of Xylulose-5-Phosphate
        - Activation of PP2A by Xylulose-5-phosphate
      - AMPK inhibits chREBP transcriptional activation activity
      - PP2A-mediated dephosphorylation of key metabolic factors
      - ChREBP activates metabolic gene expression
    - Metabolism of nitric oxide: NOS3 activation and regulation (1/41) FDR: 6.9E-1
    - The citric acid (TCA) cycle and respiratory electron transport (11/235) FDR: 6.62E-1
    - Metabolism of nucleotides (3/253) FDR: 9.76E-1
    - Metabolism of vitamins and cofactors (9/377) FDR: 7.48E-1
    - Metabolism of amino acids and derivatives (15/661) FDR: 8.48E-1
    - Metabolism of porphyrins (2/74) FDR: 6.62E-1
    - Biological oxidations (9/545) FDR: 9.74E-1





## Input data

Expand a gene, a term, or a variant into a gene set:

e.g. STAT3, breast cancer, or rs28897756



Try an example

Include the top 100 most relevant genes



Ген, термин, вариант → список генов

- коэкспрессия
- встречаемость в литературе
- и др

Посмотреть огромный список доступных аннотаций

Paste a set of Entrez gene symbols on each row in the textbox below. You can try a gene set [example](#). Also, you can now try adding a [background](#).

Paste a set of valid Entrez gene symbols (e.g. STAT3) on each row in the text-box

Скопировать свой список генов сюда  
! Нужны символьные названия генов !

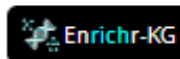
0 gene(s) entered

In order to enable others to search your set please enter a brief description of it.

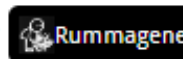
Contribute your set so it can be searched by others

Submit

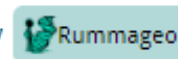
Try



Try



Try





## Input data

Expand a gene, a term, or a variant into a gene set:



Try an example

Include the top 100 most relevant genes



1. Задали список  
интересных генов:

Paste a set of Entrez gene symbols on each row in the textbox below. You can try a gene set [example](#) ([clear](#)). Also, you can now try adding a [background](#).

```
Nsun3
Polrmt
Nlrp1
Sfxn5
Zc3h12c
Slc25a39
Arsg
Defb29
Ndufb6
Zfand1
```

375 gene(s) entered

 Contribute your set so it can be searched by others

2. Жмякнули кнопку



# Enrichr

## Группы аннотаций

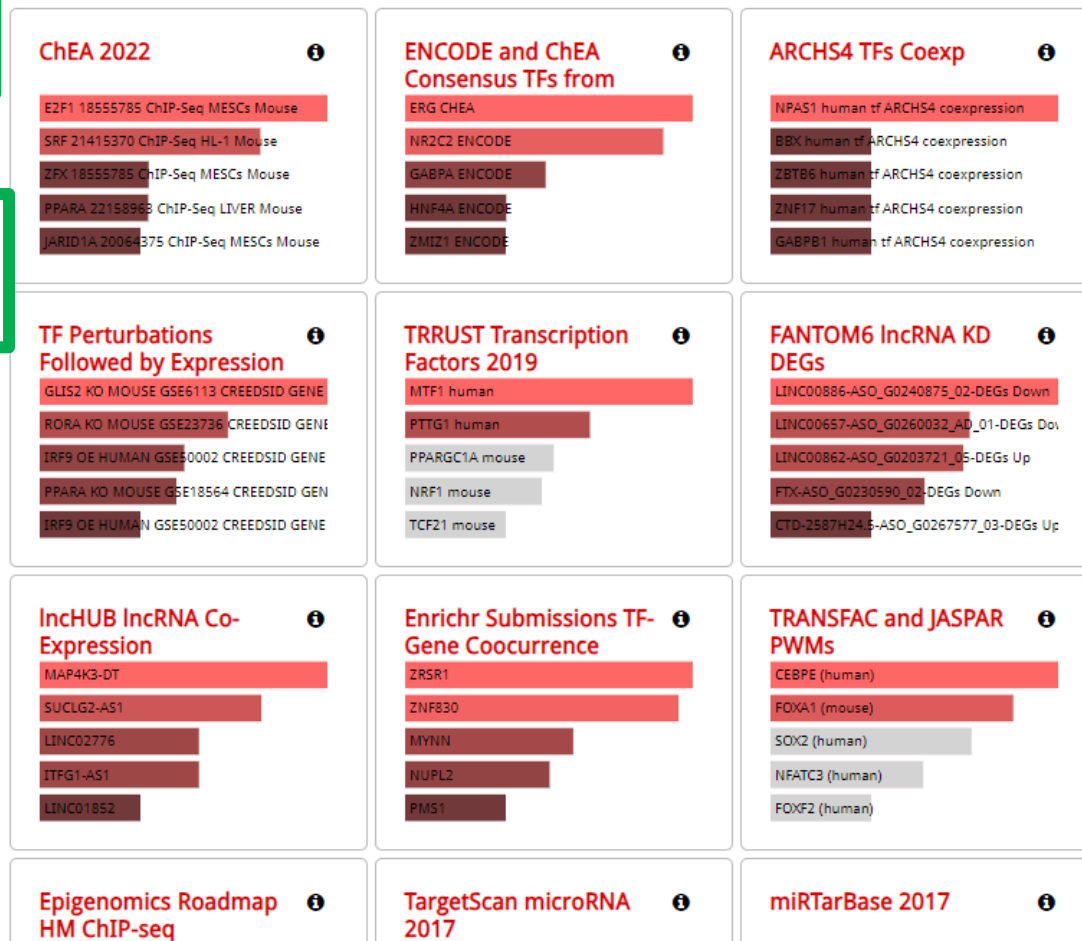
[Login](#) | [Register](#)

**Transcription** Pathways Ontologies Diseases/Drugs Cell Types Misc Legacy Crowd

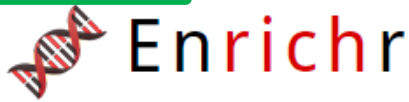
**Description** Sample gene set (375 genes)

Получили результат - радуемся!

Еще больше результатов по разным аннотациям



Изучили результаты по нужным аннотациям!



Выбрали группу аннотаций

Login | Register

Transcription Pathways Ontologies **Diseases/Drugs** Cell Types Misc Legacy Crowd

Description Sample gene set (375 genes)

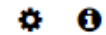
### COVID-19 Related Gene Sets 2021

**Bar Graph**

Table

Clustergram

Appyter



Click the bars to sort. Now sorted by p-value ranking.

SVG PNG JPG

- 500 genes down-regulated by MHV-A59 in murine liver cells from GSE146074 5d
- 447 genes down-regulated by SARS-CoV-2 infection in Vero E6 cells from GSE153940
- Top 500 up genes for SARS-CoV-2 infection in Rhesus macaques at Group 2 dose in PBMCs at 10 DPI from GSE15
- 500 genes down-regulated by SARS-CoV-2 in A549-ACE2 cells from GSE154613 terfenadine
- Top 500 down genes for SARS-CoV-2 infection in Rhesus macaques at Group 1 dose in PBMCs at 4 DPI from GSE1
- COVID19-Nsp13 protein host PPI from Krogan
- 500 genes down-regulated by SARS-CoV-2 in mouse Kidney cells at 7 dpi from GSE162113
- Top 500 downregulated genes in mouse kidney with SARS-CoV-2 infection (Day 7) from GEO GSE162113
- 332 proteins enriched in AP-MS using various SARS-CoV-2 proteins as bait from Gordon et al. (2020)
- COVID19-All 332 protein host PPI from Krogan

Обогащенные категории

Выбрали вид представления результатов

MSigDB Home

Human Collections

- ▶ About
- ▶ Browse
- ▶ Search
- ▶ Investigate
- ▶ Gene Families

Mouse Collections

- ▶ About
- ▶ Browse
- ▶ Search
- ▶ Investigate

Help



**MSigDB**  
Molecular Signatures  
Database

## Molecular Signatures Database

### Overview

The Molecular Signatures Database (MSigDB) is a resource of tens of thousands of annotated gene sets for use with GSEA software, divided into Human and Mouse collections. From this web site, you can

- ▶ **Examine** a gene set and its annotations. See, for example, the [HALLMARK\\_APOPTOSIS human gene set page](#).
- ▶ **Browse** gene sets by name or collection.
- ▶ **Search** for gene sets by keyword.
- ▶ **Investigate** gene sets:
  - ▶ **Compute overlaps** between your gene set and gene sets in MSigDB.
  - ▶ **Categorize** members of a gene set by gene families.
  - ▶ **View the expression profile** of a gene set in a provided public expression compendia.
  - ▶ Investigate the gene set in the online **biological network repository NDEX**
- ▶ **Download** gene sets.

### License Terms

GSEA and MSigDB are available for use under [these license terms](#).

Please [register](#) to download the GSEA software and the MSigDB gene sets, and to use our web tools. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

### Current Version

Human MSigDB v2023.2.Hs updated October 2023. [Release notes](#).

Mouse MSigDB v2023.2.Mm updated October 2023. [Release notes](#).

### Human Collections

<b>H</b> <b>hallmark gene sets</b> are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.	<b>C5</b> <b>ontology gene sets</b> consist of genes annotated by the same ontology term.
<b>C1</b> <b>positional gene sets</b> corresponding to human chromosome cytogenetic bands.	<b>C6</b> <b>oncogenic signature gene sets</b> defined directly from microarray gene expression data from cancer gene perturbations.
<b>C2</b> <b>curated gene sets</b> from online pathway databases, publications in PubMed, and knowledge of domain experts.	<b>C7</b> <b>immunologic signature gene sets</b> represent cell states and perturbations within the immune system.
<b>C3</b> <b>regulatory target gene sets</b> based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.	<b>C8</b> <b>cell type signature gene sets</b> curated from cluster markers identified in single-cell sequencing studies of human tissue.
<b>C4</b> <b>computational gene sets</b> defined by mining large collections of cancer-oriented expression data.	

### Mouse Collections

<b>MH</b> <b>mouse-ortholog hallmark gene sets</b> are versions of gene sets in the MSigDB Hallmarks collection mapped to their mouse orthologs.	<b>M3</b> <b>regulatory target gene sets</b> based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.
<b>M1</b> <b>positional gene sets</b> corresponding to mouse chromosome cytogenetic bands.	<b>M5</b> <b>ontology gene sets</b> consist of genes annotated by the same ontology term.

Аннотированные гены человека и и мыши сгруппированы по 9 и 6 коллекциям

Можно скачать все сразу, либо отдельные коллекции или аннотации внутри коллекций.



Коллекции включают в себя и широко известные GO, KEGG,

Reactome...

так и более

специализированные

коллекции:

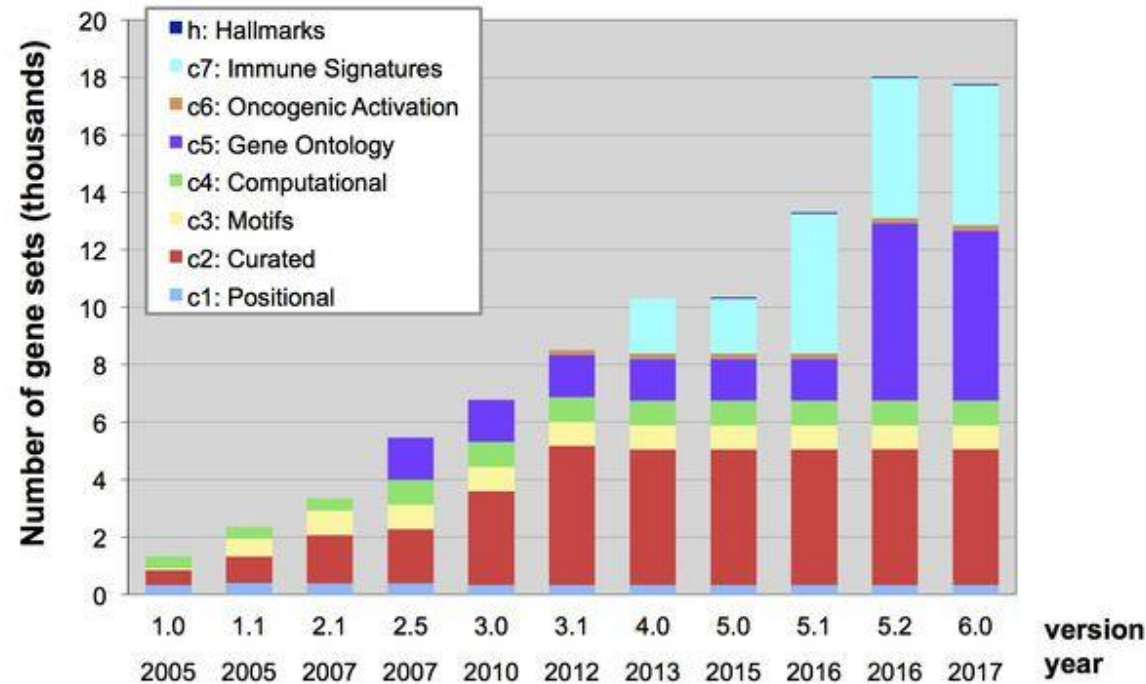
- связанные с иммунной системой
  - раком
  - вакцинами...

Есть и маркеры клеточных типов, собранные по нескольким scRNA-seq публикациям.

На этом же сайте можно скачать ПО для GSEA анализа.

Так же можно использовать веб-сервис

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



UNIVERSITY OF CALIFORNIA  
**SANTA CRUZ** Genomics Institute



# Genome Browser

UCSC Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us



Search genes, data, help docs and more...

Search

## Tools



hg38



hg19



mm39

- **Genome Browser** - Interactively visualize genomic data
- **BLAT** - Rapidly align sequences to the genome
- **In-Silico PCR** - Rapidly align PCR primer pairs to the genome
- **Table Browser** - Download and filter data from the Genome Browser
- **LiftOver** - Convert genome coordinates between assemblies
- **REST API** - Returns data requested in JSON format
- **Variant Annotation Integrator** - Annotate genomic variants
- **More tools...**

## News

- Mar. 26, 2024 - **New gnomAD v4 Constraint Metric**
- Mar. 07, 2024 - **New Prediction Scores super track**
- Mar. 05, 2024 - **New JASPAR tracks: Human (hg19)**
- Mar. 01, 2024 - **AbSplice Prediction Scores for hg19**
- Feb. 21, 2024 - **New DECIPHER Dosage Sensitivity**
- Feb. 14 2024 - **New GENCODE gene tracks: V45 (1)**

More news...

Subscribe

## Browse/Select Species

### POPULAR SPECIES



Enter species or common name

### REPRESENTED SPECIES

- Human
- Chimp
- Bonobo
- Gorilla
- Orangutan
- Gibbon
- Green monkey
- Crab-eating macaque
- Rhesus
- Baboon (anubis)
- Baboon (hamadryas)
- Proboscis monkey

## Find Position

Human Assembly  
Dec. 2013 (GRCh38/hg38)

GO

### Position/Search Term

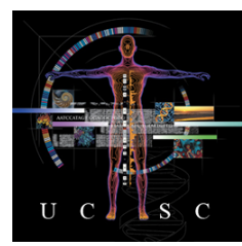
Enter position, gene symbol or search terms

Current position: chrX:15,560,138-15,602,945

## Human Genome Browser - hg38 assembly

[view sequences](#)

UCSC Genome Browser assembly ID: hg38  
Sequencing/Assembly provider ID: Genome Reference Consortium Human GRCh38.p12 (GCA\_000001405.27)  
Assembly date: Dec. 2013 initial release; Dec. 2017 patch release 12  
Assembly accession: [GCA\\_000001405.27](#)  
NCBI Genome ID: 51 (Homo sapiens (human))  
NCBI Assembly ID: [5800238](#) (GRCh38.p12, GCA\_000001405.27)  
BioProject ID: [PRJNA31257](#)



Homo sapiens  
(Graphic courtesy of CBSE)

Search the assembly:

- 1) Выбираем человека – Human
- 2) Версия генома человека – GRCh38/hg38
- 3) В раздел Position/Search Term вводим название своего гена или координаты локуса



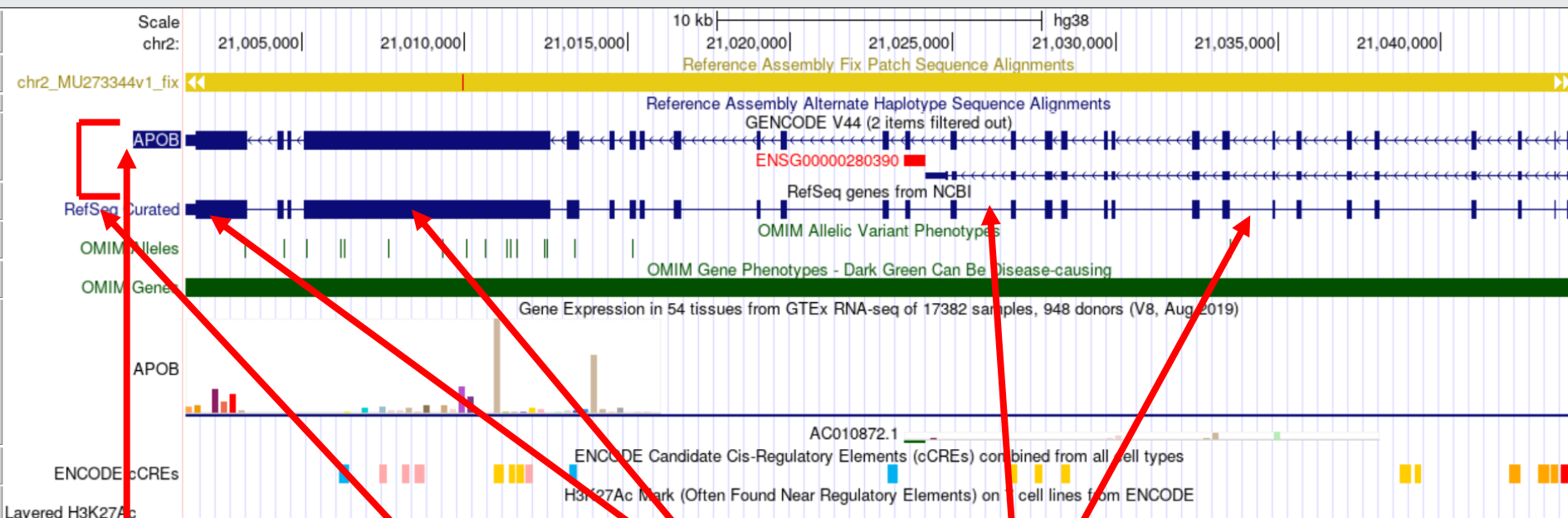
### UCSC Genome Browser on Human (GRCh38/hg38)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

A guided tutorial is available for new users: [Start tutorial](#)

[Close](#) [Don't show again](#)

multi-region chr2:21,001,429-21,044,073 42,645 bp.  gene, chromosome range, search terms, help pages, see exan [go](#) [examples](#)



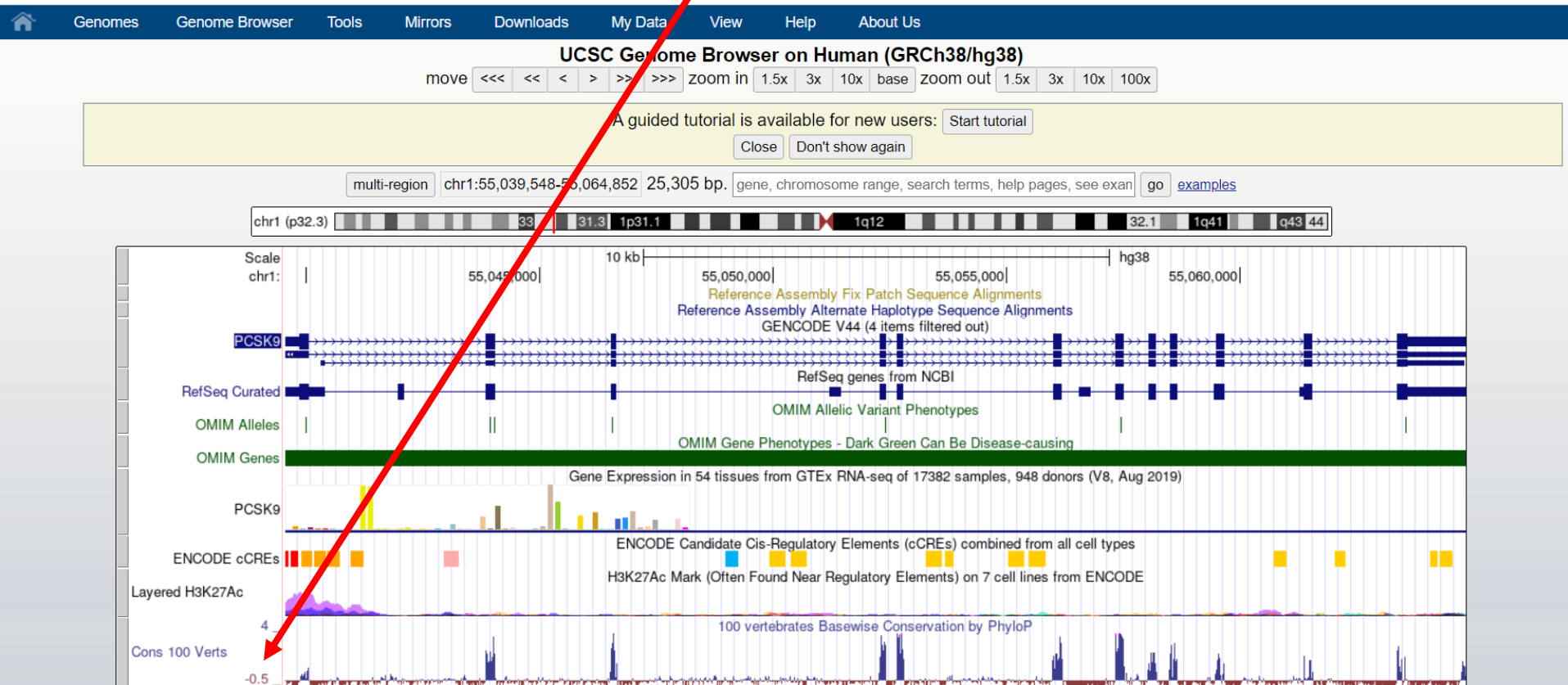
Ген

Экзоны

Интроны

Нажимаем на название гена и попадаем на страницу его подробного описания

# Обратите внимание на трек консервативности



# Можно спрятать все треки

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

chr1:55,039,548-55,064,852 25,305 bp. enter position, gene symbol, HGVS or search terms go

chr1 (p32.3) hg38

Scale chr1: 55,045,000 10 kb 55,050,000 55,055,000 55,060,000

move start < 2.0 > move end < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.

track search default tracks default order **hide all** add custom tracks track hubs configure multi-region reverse resize refresh

collapse all Use drop down controls below and press refresh to alter tracks displayed. expand all

Tracks with lots of items will automatically be displayed in more compact modes.

Mapping and Sequencing refresh

# Открыть только некоторые

hide hide hide

**Genes and Gene Predictions** refresh

P12 [GENCODE v32](#) [RefSeq](#) [Other RefSeq](#) [Updated All GENCODE...](#) [AUGUSTUS](#) [CCDS](#)

hide hide hide hide hide hide

P12 [Genscan](#) [IKMC Genes](#)

**Repeats** refresh

P12 [RepeatMasker](#) [Interrupted Rpts](#) [RepeatMasker Viz.](#) [Microsatellite](#) [Segmental Dups](#) [Self Chain](#)

full hide hide hide hide hide

P12 [Simple Repeats](#) [WM + SDust](#)

hide hide

<https://genome.ucsc.edu/>



UNIVERSITY OF CALIFORNIA  
**SANTA CRUZ** Genomics  
Institute



## Genome Browser



Genomes

Genome Browser

Tools

Mirrors

Downloads

My Data

Projects

Help

About Us



### Our tools

- **Genome Browser**  
interactively visualize genomic data
- **BLAT**  
rapidly align sequences to the genome
- **Table Browser**  
download data from the Genome Browser database
- **Variant Annotation Integrator**  
get functional effect predictions for variant calls
- **Data Integrator**  
combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBiB)**  
run the Genome Browser on your laptop or server
- **In-Silico PCR**  
rapidly align PCR primer pairs to the genome
- **LiftOver**  
convert genome coordinates between assemblies
- **Track Hubs**  
import and view external data tracks
- **REST API**  
returns data in JSON format

<https://genome.ucsc.edu/>



UNIVERSITY OF CALIFORNIA  
**SANTA CRUZ** Genomics  
Institute



## Genome Browser



Genomes

Genome Browser

**Tools**

Mirrors

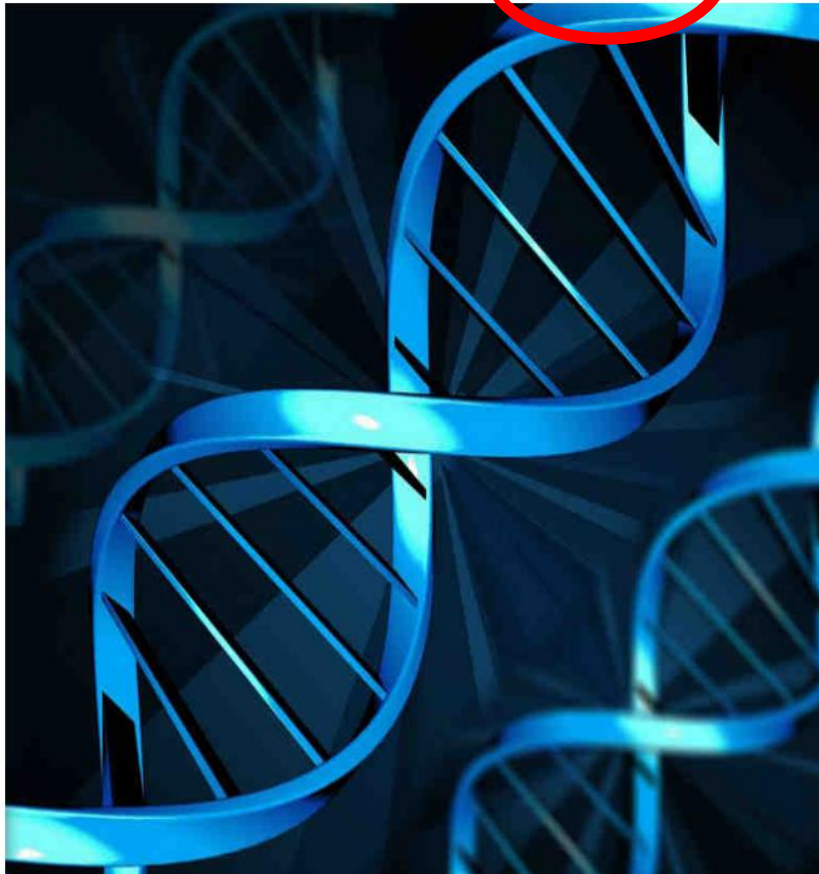
Downloads

My Data

Projects

Help

About Us



### Our tools

- **Genome Browser**  
interactively visualize genomic data
- **BLAT**  
rapidly align sequences to the genome
- **Table Browser**  
download data from the Genome Browser database
- **Variant Annotation Integrator**  
get functional effect predictions for variant calls
- **Data Integrator**  
combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBiB)**  
run the Genome Browser on your laptop or server
- **In-Silico PCR**  
rapidly align PCR primer pairs to the genome
- **LiftOver**  
convert genome coordinates between assemblies
- **Track Hubs**  
import and view external data tracks
- **REST API**  
returns data in JSON format





## Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, and the [User's Guide](#) for general information and sample queries. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

**clade:**  **genome:**  **assembly:**

**group:**  **track:**

**table:**

**region:**  genome  position

**filter:**

**subtrack merge:**

**intersection:**

**correlation:**

**output format:**  Send output to  [Galaxy](#)  [GREAT](#)

**output file:**  (leave blank to keep output in browser)

**file type returned:**  plain text  gzip compressed

*Note: to return more than 100,000 lines, change the filter setting (above). The entire data set may be available for download as a very large file that contains the original data values (not compressed into the wiggle format) -- see the Downloads page.*

To reset **all** user cart settings (including custom tracks), [click here](#).

## GeneCards®: The Human Gene Database

GeneCards is a searchable, integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. The knowledgebase automatically integrates gene-centric data from ~150 web sources, including genomic, transcriptomic, proteomic, genetic, clinical and functional information.



### Explore a Gene

IL6 [GO](#)

#### Jump to section for this gene:

- |                          |                           |                          |                          |                              |                          |                           |                              |                           |
|--------------------------|---------------------------|--------------------------|--------------------------|------------------------------|--------------------------|---------------------------|------------------------------|---------------------------|
| <a href="#">Aliases</a>  | <a href="#">Disorders</a> | <a href="#">Domains</a>  | <a href="#">Drugs</a>    | <a href="#">Expression</a>   | <a href="#">Function</a> | <a href="#">Genomics</a>  | <a href="#">Localization</a> | <a href="#">Orthologs</a> |
| <a href="#">Paralogs</a> | <a href="#">Pathways</a>  | <a href="#">Products</a> | <a href="#">Proteins</a> | <a href="#">Publications</a> | <a href="#">Sources</a>  | <a href="#">Summaries</a> | <a href="#">Transcripts</a>  | <a href="#">Variants</a>  |

### GeneCardsSuite

#### NGS Analysis



#### Affiliated Databases



#### Analysis Tools



# IL6 Gene (Protein Coding) ★

Interleukin 6

GCID: GC07P022725 <sup>?</sup>

GIFTS: 50 <sup>?</sup>



Genes  
Participants

Jump to section	Aliases Paralogs	Disorders Pathways	Domains Products	Drugs Proteins	Expression Publications	Function Sources	Genomics Summaries	Localization Transcripts	Orthologs Variants
Research Products	Antibodies Cell Lines	Assays Clones	Proteins Primers	Inhib. RNA Genotyping	CRISPR	Exp. Assays	miRNA	Drugs	Animal Models

**R&D** Proteins Primary Antibodies  
ELISAs Antibody Arrays  
Activity Assays

**ORIGENE** Proteins Antibodies Assays Genes  
shRNA Primers CRISPR  
Lentiviral Particles

**NOVUS BIOLOGICALS** Antibodies Proteins Lysates  
Elisa Kits

**GenScript** Gene Synthesis ORF clones  
Peptide Service Peptide Library  
CRISPR

## Aliases for IL6 Gene <sup>?</sup>

### Aliases for IL6 Gene

- |  |   |
|--|---|
| Interleukin 6 <sup>2 3 5</sup>             | IFNB2 <sup>3 4</sup>                            |
| IL-6 <sup>2 3 4</sup>                      | BSF2 <sup>2 3</sup>                             |
| B-Cell Stimulatory Factor 2 <sup>3 4</sup> | CDF <sup>3 4</sup>                              |
| CTL Differentiation Factor <sup>3 4</sup>  | HGF <sup>2 3</sup>                              |
| Hybridoma Growth Factor <sup>3 4</sup>     | HSF <sup>2 3</sup>                              |
| Interferon Beta-2 <sup>3 4</sup>           | Interleukin 6 (Interferon, Beta 2) <sup>2</sup> |
| Interleukin-6 <sup>3 4</sup>               | B-Cell Differentiation Factor <sup>3</sup>      |
| IFN-Beta-2 <sup>3 4</sup>                  | Interferon, Beta 2 <sup>2</sup>                 |
| BSF-2 <sup>3 4</sup>                       | Interleukin BSF-2 <sup>3</sup>                  |
|  | IL6 <sup>5</sup>                                |

## GeneCards for AI/ML <sup>X</sup>

Access comprehensive integrated data from >190 biomedical sources in machine-readable formats

JSON  
XML  
CSV  
API  
...

AI >



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



BLAST Align Peptide search **ID mapping** SPARQL

Release 2024\_02 | Statistics

# Find your protein

UniProtKB ▾

Advanced | List Search

Examples: Insulin, APP, Human, P05067, organism\_id:9606

UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. [Cite UniProt](#)

UniProt invites you to participate in a survey on the use and value of UniProt. Please click on the link provided to complete the survey .

Proteins

Species

Protein Clusters

Sequence Archive

## Retrieve/ID mapping

Enter one or more IDs (100,000 max). You may also [load from a text file](#). Separate IDs by whitespace (space, tab, newline) or commas.

P31946

 Your input contains 1 ID

From database

UniProtKB AC/ID ▾

To database

Gene Name ▾



Указываем базы данных

# ID mapping 1 result found for UniProtKB\_AC-ID → Gene\_Name

[Overview](#) [Input Parameters](#) [API Request](#)

[Download](#) View:  Cards  Table [Resubmit](#)

1 ID was mapped to 1 result

From	To
P31946	YWHAB

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

1. Enter ids and or select file for batch upload. Else enter ids or select file or list from workspace for comparing to a reference list.

Enter IDs:   
[Supported IDs](#)  separate IDs by a space or comma

Upload IDs:  Файл не выбран  
[File format](#)

Please [login](#) to be able to select lists from your workspace.

Select List Type:  ID List  
 Previously exported text search results  
 Workspace list  
 PANTHER Generic Mapping  
 ID's from Reference Proteome Genome  
Organism for id list   
 VCF File Flanking region   Search Enhancer Data

2. Select organism.  
  
Mus musculus  
Rattus norvegicus  
Gallus gallus  
Danio rerio

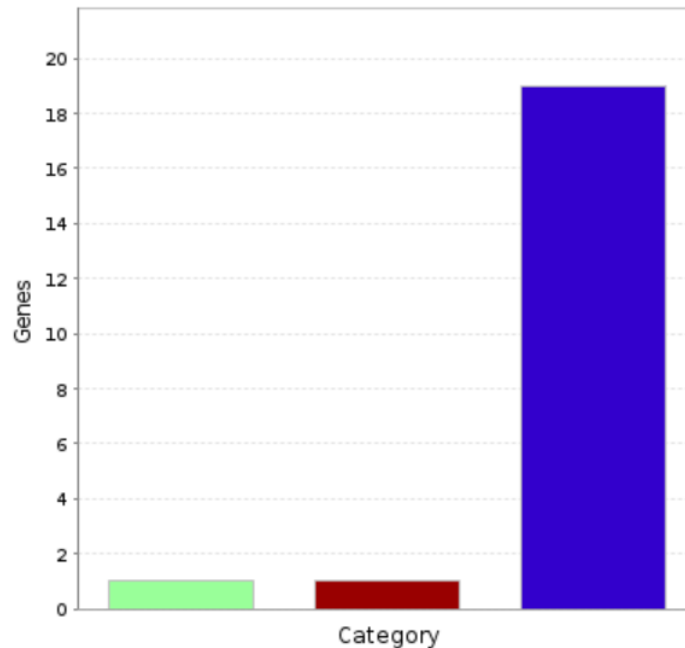
3. Select Analysis.  
 Functional classification viewed in gene list  
 Functional classification viewed in graphic charts  Bar chart  Pie chart

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

Select Ontology:  View:

**PANTHER Protein Class**

Total # Genes: 22 Total # protein class hits: 21



Click to get gene list for a category:

- [RNA metabolism protein \(PC00031\)](#)
- [protein-binding activity modulator \(PC00095\)](#)
- [translational protein \(PC00263\)](#)

Color picker powered by  Web Colors by VistBone