

# Лекция №5

## **Интеграция разрозненных источников информации по биомедицинской тематике**

м.н.с. лаб. Компьютерной протеомики Сайк Ольга Владимировна

# План лекции

1. Введение.
2. Обзор Интернет-ресурсов, содержащих разнообразную информацию об участниках генных сетей, представленную в различных форматах.
3. Интеграция разрозненных источников информации (единый формат, единый способ графического и текстового представления, возможность выполнения поискового запроса ко всем данным). Преимущества интеграция биологических данных.
4. Ресурсы, интегрирующие биологическую информацию из разнородных источников и представляющие ее в виде генных сетей: ANDSystem, STRING, GeneMania, Pathway Commons.
5. Примеры исследований, выполненных с использованием этих ресурсов.

# 1. Введение

**Генная сеть** – группа координированно функционирующих генов, взаимодействующих друг с другом как через свои первичные продукты (РНК и белки), так и через разнообразные метаболиты и другие вторичные продукты функционирования генных сетей, которая контролирует какой-либо фенотипический признак организма.

## **Компоненты генной сети:**

- 1) группа координированно экспрессирующихся генов, составляющая ядро сети;
- 2) белки, кодируемые этими генами;
- 3) низкомолекулярные компоненты (гормоны и другие сигнальные молекулы, энергетические компоненты, метаболиты);
- 4) связи между участниками сети (в том числе отрицательные и положительные обратные связи).

A gene network is a mixed graph  $G : = (V, U, D)$  over a set  $V$  of nodes, corresponding to gene activities, with unordered pairs  $U$ , the undirected edges, and ordered pairs  $D$ , the directed edges.

Pinna A., Soranzo N., De La Fuente A. From knockouts to networks: establishing direct cause-effect relationships through graph analysis // PloS one. – 2010. – T. 5. – №. 10. – C. e12912.

A gene network is a directed labelled graph, where each node represents a gene and each arc represents a relation between the genes.

Rung J. et al. Building and analysing genome-wide gene disruption networks // Bioinformatics. – 2002. – T. 18. – №. suppl 2. – C. S202-S210.

Gene network is a graphical illustration for exploring the functional linkages and the potential coordinate regulations of genes.

Wang et al. Gene Network Exploration of Crosstalk between Apoptosis and Autophagy in Chronic Myelogenous Leukemia // BioMed Research International, 2014.

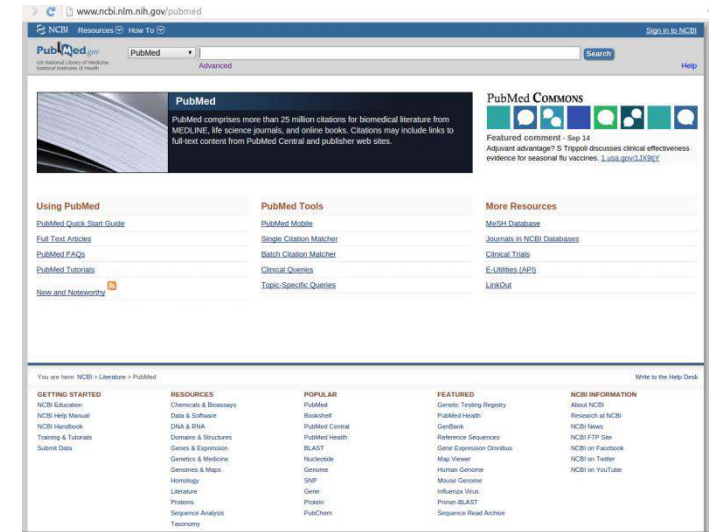
**A gene network is a collection of effective interactions, describing the multiple ways through which one gene affects all the others to which it is connected.**

Zhu Y., Pan W., Shen X. Support vector machines with disease-gene-centric network penalty for high dimensional microarray data // Statistics and its interface. – 2009. – T. 2. – №. 3. – C. 257.

# Источники информации

- Важнейшим источником информации по биомедицинской тематике являются научные публикации. В системе **PubMed**, которая является крупнейшей базой данных научных статей, доступных через интернет, собрано более 26 миллионов публикаций.
- Базы данных. В журнале **Nucleic Acids Research (NAR) database issue** (импакт фактор 8,65) опубликованы статьи по более чем 1900 различным базам данных, содержащих информацию по биомедицинской тематике. Каждая база данных содержит структурированную и формализованную информацию. Однако, формат представления информации в каждой базе данных свой!

## Главная страница системы PubMed



## Некоторые базы данных по участникам генных сетей



Как Вы думаете, по какому запросу к PubMed будет найдено больше статей?

1. *Beta vulgaris* - сахарная свёкла



2. *Panax ginseng* – женьшень обыкновенный



# Как Вы думаете, по какому запросу к PubMed будет найдено больше статей?

1. *Beta vulgaris* - сахарная свёкла:  
**3864 документа**

2. *Panax ginseng* – женьшень обыкновенный:  
**5999 документов**



NCBI Resources How To Sign in to NCBI

PubMed *Beta vulgaris* Search

US National Library of Medicine National Institutes of Health

Format: Summary Sort by: Most Recent Per page: 20

Send to Filters: Manage Filters

Search results

Items: 1 to 20 of 3864

1. [Seed-specific expression and analysis of recombinant anti-HER2 single-chain variable fragment \(scFv-Fc\) in Arabidopsis thaliana.](#)  
Dong Y, Jian L, Yao N, Wang D, Liu X, Wang N, Li X, Wang F, Li H, Jiang C.  
Protein Expr Purif. 2017 Mar 9. pii: S1046-5928(16)30220-0. doi: 10.1016/j.pep.2017.03.009. [Epub ahead of print]  
PMID: 28286176  
[Similar articles](#)

Results by year

Download CSV

Related searches

*beta vulgaris* antioxidant

NCBI Resources How To Sign in to NCBI

PubMed *Panax ginseng* Search

US National Library of Medicine National Institutes of Health

Format: Summary Sort by: Most Recent Per page: 20

Send to Filters: Manage Filters

Search results

Items: 1 to 20 of 5999

1. [The Protective Effect of the Active Components of ERPC on Diabetic Peripheral Neuropathy in Rats.](#)  
Hao GM, Liu YG, Wu Y, Xing W, Guo SZ, Wang Y, Wang ZL, Li C, Lv TT, Wang HL, Shi TJ, Wang W, Han J.  
J Ethnopharmacol. 2017 Mar 15. pii: S0378-8741(16)31590-2. doi: 10.1016/j.jep.2017.03.015. [Epub ahead of print]  
PMID: 28315720  
[Similar articles](#)

Results by year

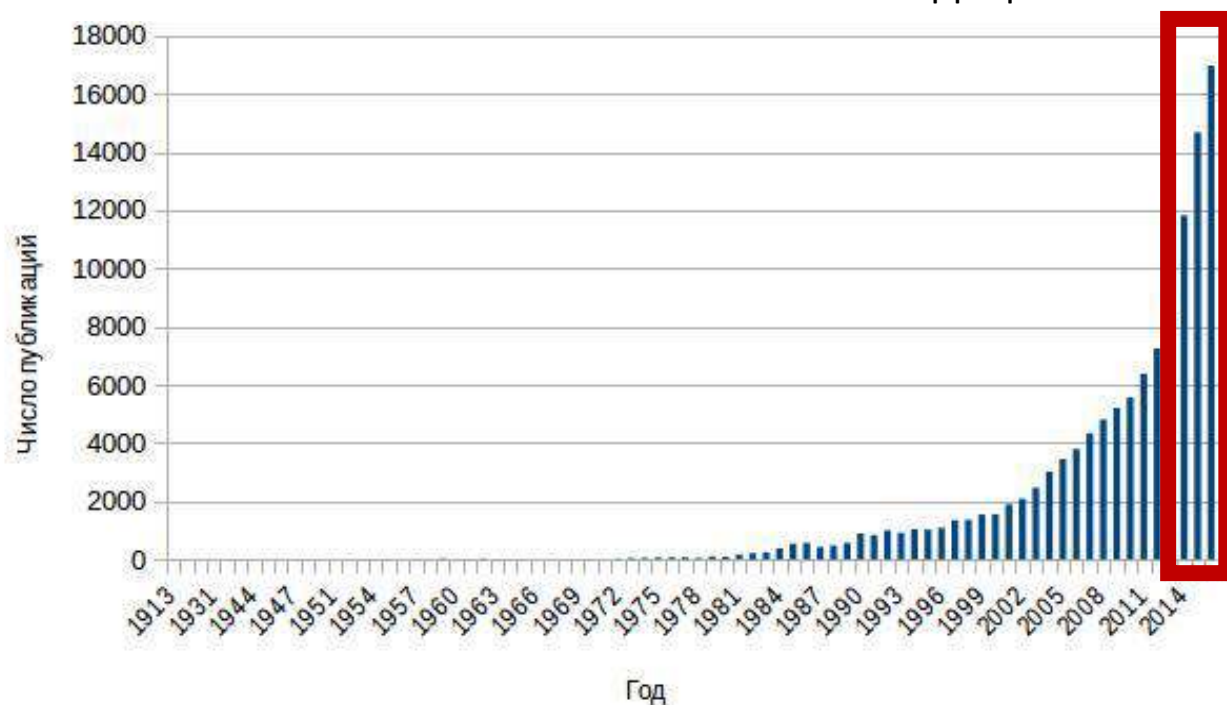
Download CSV

Related searches

*panax ginseng* review

**Женьшень** (от кит. trad. 人參) («Корень жизни») — многолетнее травянистое растение, род семейства Аралиевые. Включает 12 видов, произрастающих в Азии и Северной Америке. Хорошо известное лекарственное растение. В основном используется как адаптоген и в качестве общетонизирующего средства. В Корее и Китае корень женьшеня также используют в приготовлении пищи. Традиционная китайская медицина утверждает, что препараты женьшеня продлевают жизнь и молодость

Число документов в PubMed по запросу «chinese medicine» - китайская медицина.



Интерес к китайской народной медицине растет в последние годы. В 2014 году виден резкий скачок числа научных публикаций по этой тематике.



На «круглом столе» в Госдуме предложено усилить правовое регулирование традиционной, народной и восточной медицины в России

Комитет по охране здоровья провел «круглый стол» на тему: «Правовое регулирование и перспектива развития традиционной, народной и восточной медицины в РФ»

Политика 23 февраля 2014, 03:11 1462 Редакция INFOPOL.RU

### Китайская традиционная медицина – альтернатива западным методам

11.12.2016 37 1900

Интерес к традиционной китайской медицине растет. Она выступает в качестве альтернативы западным методам в профилактике болезней и лечении хронических заболеваний. Об этом заявила Александра Чумбаева, заместитель главного врача одной из клиник традиционной китайской медицины в Москве.





2. Обзор Интернет-ресурсов, содержащих разнообразную информацию об участниках генных сетей, представленную в различных форматах.

### **Компоненты генной сети:**

- 1) группа координированно экспрессирующихся генов, составляющая ядро сети;
- 2) белки, кодируемые этими генами;
- 3) низкомолекулярные компоненты (гормоны и другие сигнальные молекулы, энергетические компоненты, метаболиты);
- 4) связи между участниками сети (в том числе отрицательные и положительные обратные связи).

По Колчанову Н.А. с соавт. ГЕННЫЕ СЕТИ // Вавиловский журнал генетики и селекции, 2013, Т. 17:4/2.

5) И другие биологические объекты

# Базы данных, содержащие информацию по генам.



- HGNC
- NCBI Gene (Entrez Gene)
- NCBI Nucleotide (GenBank)
- GeneCards
- Ensembl
- UCSC Genome Browser
- GEO – хранилище данных по экспрессии генов
- другие

# HGNC - HUGO Gene Nomenclature Committee

- Доступна через интернет по адресу <http://www.genenames.org/>
- Содержит верифицированные названия и синонимы человеческих генов.
- Идентификаторами являются числа. Например, ген YY1 имеет идентификатор 12856.
- Содержится информация по
  - ✓ Синонимам
  - ✓ Хромосомному локусу
  - ✓ Ссылки на другие ресурсы

The screenshot displays the HGNC website interface. At the top, the URL is [www.genenames.org/cgi-bin/gene\\_symbol\\_report?hgnc\\_id=HGNC:12856](http://www.genenames.org/cgi-bin/gene_symbol_report?hgnc_id=HGNC:12856). The page title is "Symbol Report: YY1". The main content area is divided into several sections:

- APPROVED SYMBOL:** YY1
- APPROVED NAME:** YY1 transcription factor
- HGNC ID:** HGNC:12856
- PREVIOUS SYMBOLS & NAMES:** -
- SYNONYMS:** DELTA, "INO80 complex subunit S", INO80S, NF-E1, UCRBP, "Yin and Yang 1 protein", YIN-YANG-1
- LOCUS TYPE:** gene with protein product
- CHROMOSOMAL LOCATION:** 14q32.2
- GENE FAMILY:** INO80 complex  
Zinc fingers C2H2-type
- HCOP:** Orthology Predictions for YY1

Below the main report is the "External links" section, which includes a "HOMOLOGS" table and several "GENE RESOURCES" and "NUCLEOTIDE SEQUENCES" lists.

|                   | Symbol | Database                    |
|-------------------|--------|-----------------------------|
| Mus musculus      | Yy1    | <a href="#">MGI:99150</a> C |
| Rattus norvegicus | Yy1    | <a href="#">RGD:3982</a> D  |

**GENE RESOURCES**

- Entrez Gene: [7528](#) C
- Ensembl: [ENSG00000100811](#) C [Region in detail](#) [Sequence](#)
- Vega: [QTHUMG00000150479](#) C [Region in detail](#) [Sequence](#)
- UCSC: [uc001ygy.3](#) D [Genome browser](#)

**NUCLEOTIDE SEQUENCES**

- BC020324 C [GenBank](#) [ENA](#) [DDBJ](#)
- NM\_003403 C [RefSeq](#) [NCBI Sequence Viewer](#)
- CCDS9957 C [CCDS](#)

**PROTEIN RESOURCES**

- P25490 D [UniProt](#) [InterPro](#) [PDBe](#)

**CLINICAL RESOURCES**

- OMIM: [600013](#) D
- [GeneTests](#) D
- [Orphanet](#) D
- [DECIPHER](#) D

# NCBI Gene

- Доступна через интернет по адресу <https://www.ncbi.nlm.nih.gov/gene>
- Интегрирует информацию по генам 9479 эукариот, 1151 бактерий, 7152 вирусов и др.
- Вся информация может быть скачана через ftp сайт.
- Идентификаторами генов являются числа. Например, ген транскрипционного фактора YY1 человека имеет идентификатор 7528, мыши – 22632, крысы – 24919, *Arabidopsis thaliana* – 826093.

Поиск

Окно для ввода

запросов

Сохранение

Организмы

| Name/Gene ID      | Description   | Location   | Aliases                                 |
|-------------------|---|--|---|
| YY1<br>ID: 7528   | YY1 transcription factor [ <i>Homo sapiens</i> (human)]                             | Chromosome 14, NC_000014.9 (100238765..100279034)        | DELTA, INO80S, NF-E1, UCRBP, YIN-YANG-1 |
| Yy1<br>ID: 22632  | YY1 transcription factor [ <i>Mus musculus</i> (house mouse)]                       | Chromosome 12, NC_000078.6 (108793269..108816632)        | AW488674, NF-E1                         |
| Yy1<br>ID: 24919  | YY1 transcription factor [ <i>Rattus norvegicus</i> (Norway rat)]                   | Chromosome 6, NC_005105.4 (132702443..132726848)         | NF-E1, NMP-1, NMP1, UCRBP               |
| YY1<br>ID: 826093 | zinc finger (C2H2 type) family protein [ <i>Arabidopsis thaliana</i> (thale cress)] | Chromosome 4, NC_003075.7 (3764241..3766671, complement) | AT4G06634, AtYY1, Yin Yang 1            |
| YY1<br>ID: 534353 | YY1 transcription factor [ <i>Arabidopsis thaliana</i> (thale cress)]               | Chromosome 21, AC_000178.1 (60017711..60017751)          | AtYY1, Yin Yang 1                       |

# NCBI Gene. Информационная карточка гена.

Идентификатор

## Общая информация

- Символ гена
- Название гена
- Ссылки на другие базы данных
- Организм
- Таксономическое положение
- Синонимы
- Описание гена
- Ортологи

Full Report ▾ Send to: ▾ Hide sidebar >>

**YY1 YY1 transcription factor [ *Homo sapiens* (human) ]**  
Gene ID: 7528, updated on 12-Mar-2017

**Summary**

**Official Symbol** YY1 provided by [HGNC](#)  
**Official Full Name** YY1 transcription factor provided by [HGNC](#)  
**Primary source** [HGNC:HGNC:12856](#)  
**See related** [Ensembl:ENSG00000100811](#) [MIM:600013](#); [Vega:OTTHUMG00000150479](#)  
**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** DELTA; NF-E1; UCRBP; INO80S; YIN-YANG-1  
**Summary** YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins. The protein is involved in repressing and activating a diverse number of promoters. YY1 may direct histone deacetylases and histone acetyltransferases to a promoter in order to activate or repress the promoter, thus implicating histone modification in the function of YY1. [provided by RefSeq, Jul 2008]  
**Orthologs** [mouse](#) [all](#)

**Genomic context**

Location: 14q32.2 See YY1 in [Genome Data Viewer](#) [Map Viewer](#)

Exon count: 5

| Annotation release  | Status  | Assembly  | Chr | Location    |
|---------------------|---------|-----------|-----|-------------|
| <a href="#">108</a> | current | GRCh38.p7 | 14  | NC_000014.9 |

**Table of contents**

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

**Genome Browsers**

- Genome Data Viewer
- Map Viewer
- Variation Viewer (GRCh37.p13)
- Variation Viewer (GRCh38)

## Более детальная информация

- Расположение на хромосоме
- Интрон/экзонная структура
- Полиморфизмы
- Литература и др.

# NCBI Nucleotide

- Доступна через интернет по адресу <https://www.ncbi.nlm.nih.gov/nucleotide>
- Коллекция нуклеотидных последовательностей из нескольких источников GenBank, RefSeq, TPA и PDB.
- Информация может быть скачана через ftp сайт.
- Идентификаторы обычно начинаются с двух заглавных букв, далее нижнее подчеркивание и число. Например, одна из последовательностей гена транскрипционного фактора YY1 человека имеет идентификатор NM\_003403.4, другая - NG\_046908.1, мыши – NM\_009537.3

The screenshot shows the NCBI Nucleotide search interface. The search term 'yy1' is entered in the search box. The results page displays a list of 3830 nucleotide sequences. The first few results are:

- Synthetic construct Homo sapiens clone ccsbBroadEn\_07141 YY1 gene, encodes complete protein**  
1,374 bp linear other-genetic  
Accession: KJ897747.1 GI: 649121807  
[GenBank](#) [FASTA](#) [Graphics](#)
- Homo sapiens YY1 transcription factor (YY1), mRNA**  
3,159 bp linear mRNA  
Accession: NM\_003403.4 GI: 459683878  
[GenBank](#) [FASTA](#) [Graphics](#)
- Rattus norvegicus YY1 transcription factor (Yy1), mRNA**  
1,236 bp linear mRNA  
Accession: NM\_173290.1 GI: 27545349  
[GenBank](#) [FASTA](#) [Graphics](#)
- Mus musculus YY1 transcription factor (Yy1), mRNA**  
2,324 bp linear mRNA  
Accession: NM\_009537.3 GI: 118130338  
[GenBank](#) [FASTA](#) [Graphics](#)
- Danio rerio YY1 transcription factor a (yy1a), mRNA**  
2,903 bp linear mRNA  
Accession: NM\_212617.1 GI: 47086800  
[GenBank](#) [FASTA](#) [Graphics](#)
- Xenopus laevis YY1 transcription factor L homeolog (yy1.L), mRNA**  
1,924 bp linear mRNA  
Accession: NM\_001087615.1 GI: 148234201  
[GenBank](#) [FASTA](#) [Graphics](#)
- PREDICTED: Canis lupus familiaris YY1 transcription factor (YY1), mRNA**  
2,220 bp linear mRNA  
Accession: XM\_001100000.1 GI: 148234201  
[GenBank](#) [FASTA](#) [Graphics](#)

The interface also includes filters for species, molecule types, source databases, genetic compartments, sequence length, release date, and revision date. The 'Results by taxon' section shows top organisms like Homo sapiens (188), Mus musculus (103), and Danio rerio (40). The 'Find related data' section allows searching for related sequences in various databases. The 'Search details' section shows the search criteria: yy1[All Fields]. The 'Recent activity' section shows recent searches, including yy1 (3095) and YY1 YY1 transcription factor [Homo sapiens].

# NCBI Nucleotide. Карточка последовательности гена.

## Особенности первичной структуры последовательности

NCBI Reference Sequence: NM\_003403.4

**Homo sapiens YY1 transcription factor (YY1), mRNA**

NCBI Reference Sequence: NM\_003403.4

FASTA Graphics

Go to: [ ]

LOCUS NM\_003403 3159 bp mRNA linear PRI 07-OCT-2016

DEFINITION Homo sapiens YY1 transcription factor (YY1), mRNA.

ACCESSION NM\_003403

VERSION NM\_003403.4

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

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

REFERENCE 1 (bases 1 to 3159)

AUTHORS Tsang DP, Wu WK, Kang W, Lee YY, Wu F, Yu Z, Xiong L, Chan AW, Tong JH, Yang W, Li MS, Lau SS, Li X, Lee SD, Yang Y, Lai PB, Yu DY, Xu G, Lo KW, Chan MT, Wang H, Lee TL, Yu J, Wong N, Yip KY, To KF and Cheng AS.

TITLE Yin Yang 1-mediated epigenetic silencing of tumour-suppressive microRNAs activates nuclear factor-kappaB in hepatocellular carcinoma

JOURNAL J. Pathol. 238 (5), 651-664 (2016)

PUBMED 26880240

REMARK GeneRIF: YY1 overexpression contributes to EZH2 recruitment for H3K27me3-mediated silencing of tumour-suppressive microRNAs, thereby activating NF-kappaB signalling in hepatocarcinogenesis.

REFERENCE 2 (bases 1 to 3159)

AUTHORS Nieborak A and Gorecki A.

TITLE Significance of the pathogenic mutation T372R in the Yin Yang 1 protein interaction with DNA--thermodynamic studies

JOURNAL FEBS Lett. 590 (6), 838-847 (2016)

PUBMED 26910132

REMARK GeneRIF: Significance of the pathogenic mutation T372R in the Yin Yang 1 protein interaction with DNA--thermodynamic studies: the mutation does not affect the secondary structure of either zinc

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the YY1 gene

A High-Density Map for Navigating the Human Polycomb Complex [Cell Rep. 2016]

Significance of the pathogenic mutation T372R in the Yin Yang 1 p [FEBS Lett. 2016]

Regulation of Transcription Factor Yin Yang 1 by SET7/9-mediated Lysin [Sci Rep. 2016]

See all...

Pathways for the YY1 gene

UCH proteinases

Deubiquitination

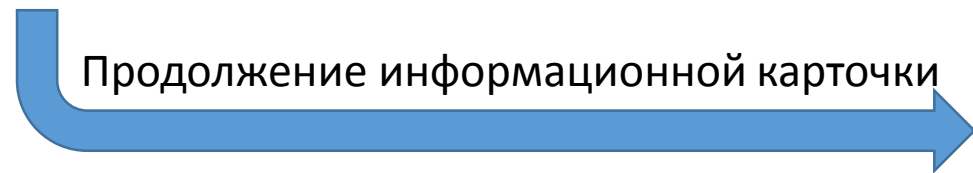
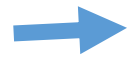
TFAP2 (AP-2) family regulates transcription of growth factors and their receptors

See all...

```
exon      1384..1542
          /gene="YY1"
          /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
          /inference="alignment:Splign:2.0.8"
exon      1543..3159
          /gene="YY1"
          /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
          /inference="alignment:Splign:2.0.8"
STS       1729..1852
          /gene="YY1"
          /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="RH1618"
          /db_xref="UniSTS:42115"
STS       1954..2640
          /gene="YY1"
          /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="YY1_3892"
          /db_xref="UniSTS:462950"
STS       2035..2173
          /gene="YY1"
          /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="G06170"
          /db_xref="UniSTS:49520"
regulatory 2690..2695
          /regulatory_class="polyA_signal_sequence"
          /gene="YY1"
          /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="RH68412"
          /db_xref="UniSTS:51218"
polyA_site 3001..3100
          /gene="YY1"
          /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="RH44921"
          /db_xref="UniSTS:36888"
regulatory 3106..3111
          /regulatory_class="polyA_signal_sequence"
          /gene="YY1"
          /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="RH68412"
          /db_xref="UniSTS:51218"
polyA_site 3132
          /gene="YY1"
          /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="RH68412"
          /db_xref="UniSTS:51218"
ORIGIN
1 agggcgaacg ggcgagtggc agcgaggcgg ggcgggctga ggccagcggc gaagtctcgc
61 gaggccggcg ccgagcagag tgtggcggcg gcggcgagat ctgggctcgg gttgaggagt
121 tggattttgt gtggaaggag gcggaaggcg aggaggaagg ggaagcggga ggcgccggcc
181 gggaggcggg aggagggcgg gccaggcggg gcggttggcg cgagcggagg cgaggcgggg
241 agcggagcgg agcagcggcg gacgagcggc gcaccgaggg cgaggaggcg
301 gggaagcccc gccgcccggc cggcgcccgc cccttcccc gccgcccggc ccctctcccc
361 ccgcccggcg gccgcccggc cccctctccc ttcttcccc acggcggcgg ccctctcccc
421 ccgcccggcg gcagccgagg agccgagggc gccgcccggc tggcggcggg gccctcagcc
```

Общая информация

Локус  
Идентификатор  
Организм и  
таксономия  
Ссылки на  
литературу



# GeneCards

- Доступна через интернет по адресу <http://www.genecards.org/>
- Интеграционный ресурс по генам ЧЕЛОВЕКА, включает информацию из более чем 125 интернет-доступных ресурсов.
- Содержится информация по
  - ✓ синонимам гена
  - ✓ Идентификаторам
  - ✓ Описание, функция
  - ✓ структуре нуклеотидной последовательности
  - ✓ продуктам гена (соответствующим белкам)
  - ✓ Биологические пути, в которых ген участвует
  - ✓ Экспрессии
  - ✓ Ортологам
  - ✓ Полиморфизмам
  - ✓ Ассоциированным заболеваниям
  - ✓ Публикациям

## Информационная карточка гена.

The screenshot displays the GeneCards website interface for the YY1 gene. The page title is "YY1 Gene (Protein Coding) ★". The gene is identified as "YY1 Transcription Factor" with a GCID of GC14P100238 and 64 GIFTs. The page includes a navigation menu with options like Home, User Guide, Analysis Tools, News And Views, and About. Below the main title, there are sections for "Aliases for YY1 Gene" listing various names like YY1 Transcription Factor, INO80 Complex Subunit S, and Delta Transcription Factor. There are also sections for "External Ids for YY1 Gene" and "Previous GeneCards Identifiers for YY1 Gene". A summary section at the bottom states: "YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins. The protein is involved in repressing and activating a diverse number of promoters. YY1 may direct histone deacetylases and histone acetyltransferases to a promoter in order to activate or repress the promoter, thus implicating histone modification in the function of YY1. [provided by RefSeq, Jul 2008]".



# Ensembl

- Доступен через интернет по адресу <http://www.ensembl.org/index.html>
- Геномный браузер для исследования геномов позвоночных (87 видов).
- Идентификаторы типа ENS код организма G/T число. Например, ген YY1 человека имеет идентификатор ENSG00000100811, один из транскриптов - ENST00000636393, ген YY1 мыши - ENSMUSG00000021264.
- Удобен для проведения сравнительных геномных исследований.
- Содержится информация по:
  - ✓ Синонимам, описанию, локализации
  - ✓ Последовательности
  - ✓ Ортологам и паралограм
  - ✓ Биологическим процессам
  - ✓ Полиморфизмам
  - ✓ Экспрессии
  - ✓ Регуляции и др.

## Информационная карточка гена.

www.ensembl.org/Homo\_sapiens/Gene/Summary?db=core;g=ENSG00000100811;r=14:100238298-100282792

Ensembl BLAST/BLAT | BioMart | Tools | Downloads | More

Human (GRCh38.p7) Location: 14:100,238,298-100,282,792 Gene: YY1

**Gene-based displays**

- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
  - Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

**Gene: YY1** ENSG00000100811

**Description** YY1 transcription factor [Source:HGNC Symbol;Acc:HGNC:12856]

**Synonyms** INO80S, hsa-mir-6764, UCRBP, YIN-YANG-1, NF-E1, DELTA

**Location** [Chromosome 14: 100,238,298-100,282,792](#) forward strand. GRCh38:CM000676.2

**About this gene** This gene has 7 transcripts ([splice variants](#)), [63 orthologues](#), [17 paralogues](#), is a member of [1 Ensembl protein family](#) and is associated with [1 phenotype](#).

**Transcripts** [Show transcript table](#)

**Summary**

**Name** [YY1](#) (HGNC Symbol)

**CCDS** This gene is a member of the Human CCDS set: [CCDS9957.1](#)

**UniProtKB** This gene has proteins that correspond to the following UniProtKB identifiers: [P25490](#)

**RefSeq** Overlapping RefSeq Gene ID [7528](#) matches and has similar biotype of protein\_coding

**Ensembl version** ENSG00000100811.11

**Other assemblies** This gene maps to [100,704,635-100,749,129](#) in GRCh37 coordinates. View this locus in the GRCh37 archive: [ENSG00000100811](#)

**Gene type** Known protein coding

**Annotation method** Annotation for this gene includes both automatic annotation from Ensembl and [Havana](#) manual curation, see [article](#).

**Alternative genes** **This gene corresponds to the following database identifiers:**  
**Havana gene:** [OTTHUMG00000150479](#)

**Annotation Attributes** overlapping locus [Definitions](#)

Go to **Region In Detail** for more tracks and navigation options (e.g. zooming)

64.50 kb Forward strand

Genes (Comprehensive set...)

YY1-004 > protein coding

YY1-002 > TEC

YY1-005 > retained intron

RP11-638I2.6-001 > lincRNA

# UCSC (University of California, Santa Cruz) Genome Browser

## Информационная карточка гена.

- Доступен через интернет по адресу <https://genome.ucsc.edu/>
- Геномный браузер для исследования геномов около 50 видов живых организмов.
- Идентификаторы типа uc три цифры три буквы . цифра. Например, один из транскриптов гена YY1 человека имеет идентификатор uc001ygy.3, другой - uc059fch.1
- Содержится информация по:
  - ✓ По интрон/экзонной структуре гена
  - ✓ Последовательности
  - ✓ Биологическим процессам
  - ✓ Полиморфизмам
  - ✓ Экспрессии
  - ✓ Регуляции и др.

The screenshot displays the UCSC Genome Browser interface for the Human Gene YY1 (ENST00000262238.8). The top navigation bar includes links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, View, and Help. The main content area is titled "UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly" and shows a genomic track for chromosome 14 (hg38) with coordinates 100,238,985-100,282,792 bp. The track displays various genomic features including RefSeq genes, OMIM Allelic Variants, H3K27ac marks, DNase I hypersensitivity peaks, and conservation scores. Below the tracks, the "Human Gene YY1 (ENST00000262238.8) Description and Page Index" is shown, including a description of the gene, a RefSeq summary, and a table of links to various databases and tools.

**Human Gene YY1 (ENST00000262238.8) Description and Page Index**

**Description:** Homo sapiens YY1 transcription factor (YY1), mRNA. (from RefSeq NM\_003403)

**RefSeq Summary (NM\_003403):** YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins. YY1 may direct histone deacetylases and histone H3 to a promoter in order to activate or repress the promoter, thus implicating histone modification in the function of YY1. [provided by RefSeq, Jul 2008]. Sequence Note: This RefSeq record was created from transcript and genomic sequence data to make the sequence consistent with the reference genome assembly. The genomic coordinates used for the transcript record were based on transcript alignments. Public RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional references. ##Evidence-Data-START## Transcript exon combination :: M77698.1, BC037308.1 [ECO:0000332] RNAseq introns :: single strand introns SAMEA1965299, SAMEA1966682 [ECO:0000348] ##Evidence-Data-END##

**GeneCode Transcript:** ENST00000262238.8

**Transcript (Including UTRs)**  
Position: hg38 chr14:100,238,985-100,282,792 Size: 43,808 Total Exon Count: 5 Strand: +

**Coding Region**  
Position: hg38 chr14:100,239,245-100,277,600 Size: 38,356 Coding Exon Count: 5

| Page Index            | Sequence and Links | UniProtKB Comments | MalaCards     | CTD            | RNA-Seq Expression |
|-----------------------|--------------------|--------------------|---------------|----------------|--------------------|
| Microarray Expression | RNA Structure      | Protein Structure  | Other Species | GO Annotations | mRNA Descriptions  |
| Pathways              | Other Names        | Methods            |               |                |                    |

Data last updated: 2016-03-28

**Sequence and Links to Tools and Databases**

| Genomic Sequence (chr14:100,238,985-100,282,792) | mRNA (may differ from genome) | Protein (414 aa)    |              |              |      |
|--|-------------------------------|---------------------|--------------|--------------|------|
| Gene Sorter                                      | Genome Browser                | Other Species FASTA | Table Schema | BioGPS       | CGAP |
| Ensembl  | Entrez Gene                   | ExonPrimer          | GeneCards    | Gepis Tissue | HGNC |
| HPRD   | Lynx                          | MGI                 | MOPED        | neXtProt     | OMIM |
| PubMed   | Reactome                      | Stanford SOURCE     | UniProtKB    | Wikipedia    |      |

**Comments and Description Text from UniProtKB**

**ID:** YY1\_HUMAN

**DESCRIPTION:** RecName: Full=Transcriptional repressor protein YY1; AltName: Full=Delta transcription factor; AltName: Full=IN subunit S; AltName: Full=NF-E1; AltName: Full=Yin and yang 1; Short=YY-1;

**FUNCTION:** Multifunctional transcription factor that exhibits positive and negative control on a large number of cellular and viral genes overlapping the transcription start site. Binds to the consensus sequence 5'-CCGCCATNTT-3'; some genes have been shown longer binding motif allowing enhanced binding; the initial CG dinucleotide can be methylated greatly reducing the binding affinity. transcription regulation is depending upon the context in which it binds and diverse mechanisms of action include direct activation

# GEO - Gene Expression Omnibus

- Доступен через интернет по адресу <https://www.ncbi.nlm.nih.gov/geo/>
- Репозиторий, в котором хранятся разнородные данные по экспрессии генов. Объектами в базе данных являются высокопроизводительные эксперименты по оценке уровней экспрессии генов (методы ДНК-микрочипов и RNA-seq).
- Идентификаторы для DataSet - GDSчисло и для DataSeries - GSEчисло. Например, серия экспериментов, посвященных анализу уровня экспрессии генов в HeLa клетках при нокауте генов YY1 и YY2, имеет идентификатор GSE14964; DataSet - GDS3788, а отдельный эксперимент этой серии - GSM373617.
- Содержится информация по:
  - ✓ Дизайну экспериментов,
  - ✓ методам, участникам, публикациям
  - ✓ Сырые и часто нормированные данные по экспрессии

## Информационная карточка эксперимента.

The screenshot shows the NCBI GEO Accession Display page for GSE14964. The page includes the NCBI logo, search options, and a navigation bar. The main content area displays the following information:

**Series GSE14964** Query DataSets for GSE14964

Status: Public on Feb 28, 2010  
Title: Genome-wide analysis of YY2 versus YY1 target genes  
Organism: Homo sapiens  
Experiment type: Expression profiling by array  
Summary: Yin Yang 1 (YY1) is a critical transcription factor controlling cell proliferation, development, and DNA damage responses. Although two homologous Drosophila YY family members (pleiohomeotic (pho)) and pleiohomeotic-like (phoI) are redundant, the functional significance of a recently described mammalian YY1-like gene (YY2) is unknown. Using microarray and gene set enrichment analysis (GSEA), we found that lentiviral constructs containing short hairpin loop YY1- and YY2-specific inhibitory RNAs (shYY1 and shYY2) caused significant changes in both redundant and distinguishable expression patterns. Ribosomal protein genes were the most significant gene set up-regulated by both shYY1 and shYY2, although combined shYY1/shYY2 knockdowns were not additive. In contrast, shYY2 reversed anti-proliferative effects of shYY1 on E2F target genes, and shYY2 particularly altered UV

Platforms (1): GPL570 [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array

Samples (13):  
GSM373614 Vector control 1  
GSM373615 Vector control 2  
GSM373616 Vector control 3  
GSM373617 First shYY1  
GSM373618 Second shYY1  
GSM373619 Third shYY1  
GSM373620 First shYY2  
GSM373621 Second shYY2  
GSM373622 Third shYY2  
GSM373623 Fourth shYY2  
GSM373624 First shYY1 and shYY2  
GSM373625 Second shYY1 and shYY2  
GSM373626 Third shYY1 and shYY2

Relations: BioProject PRJNA111841

Analyze with GEO2R

| Download family                 | Format |
|---------------------------------|--------|
| SOFT formatted family file(s)   | SOFT   |
| MINIML formatted family file(s) | MINIML |
| Series Matrix File(s)           | TXT    |

| Supplementary file | Size     | Download       | File type/resource |
|--------------------|----------|----------------|--------------------|
| GSE14964_RAW.tar   | 251.4 Mb | (http)(custom) | TAR (of CEL, CHP)  |

Raw data provided as supplementary file  
Processed data included within Sample table  
Processed data provided as supplementary file

В базе данных UCSC genome browser для человека представлено 82 960 идентификаторов и информационных карточек. Генов человека по разным оценкам от 20 до 30 тысяч.

Почему же тогда в UCSC genome browser присутствует такая цифра 82 960?

Пример информационной карточки в UCSC genome browser.

genome.ucsc.edu/cgi-bin/hgGene?db=hg19&hgg\_gene=uc003adj.3

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

### Human Gene MN1 (uc003adj.3) Description and Page Index

**Description:** Homo sapiens meningioma (disrupted in balanced translocation) 1 (MN1), mRNA.

**RefSeq Summary (NM\_002430):** Meningioma 1 (MN1) contains two sets of CAG repeats. It is disrupted by a balanced translocation (4;22) in a meningioma, and its inactivation may contribute to meningioma 32 pathogenesis. [provided by RefSeq, Jul 2008]. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications. ##Evidence-Data-START## Transcript exon combination :: X82209.2, CX866728.1 [ECO:0000332] RNAseq introns :: single sample supports all introns SAMEA2147975, SAMN03465404 [ECO:0000348] ##Evidence-Data-END##

**Transcript (Including UTRs)**  
**Position:** hg19 chr22:28,144,265-28,197,486 **Size:** 53,222 **Total Exon Count:** 2 **Strand:** -

**Coding Region**  
**Position:** hg19 chr22:28,146,903-28,196,531 **Size:** 49,629 **Coding Exon Count:** 2

|                   |                    |                       |                      |                   |               |
|-------------------|--------------------|-----------------------|----------------------|-------------------|---------------|
| <b>Page Index</b> | Sequence and Links | UniProtKB Comments    | Genetic Associations | MalaCards         | CTD           |
| Gene Alleles      | RNA-Seq Expression | Microarray Expression | RNA Structure        | Protein Structure | Other Species |
| GO Annotations    | mRNA Descriptions  | Other Names           | Model Information    | Methods           |               |

Data last updated: 2013-06-14

В базе данных UCSC genome browser для человека представлено 82 960 идентификаторов и информационных карточек. Генов человека по разным оценкам от 20 до 30 тысяч.

Почему же тогда в UCSC genome browser присутствует такая цифра 82 960?

В UCSC genome browser идентификатором «ис три цифры три буквы . цифра» (UCSC ID) обозначается один из транскриптов гена. В общем случае с гена может считываться несколько функциональных транскриптов.

genome.ucsc.edu/cgi-bin/hgGene?db=hg19&hgg\_gene=uc003adj.3

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Human Gene MN1 (uc003adj.3) Description and Page Index

**Description:** Homo sapiens meningioma (disrupted in balanced translocation) 1 (MN1), mRNA.

**RefSeq Summary (NM\_002430):** Meningioma 1 (MN1) contains two sets of CAG repeats. It is disrupted by a balanced translocation (4;22) in a meningioma, and its inactivation may contribute to meningioma pathogenesis. [provided by RefSeq, Jul 2008]. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene page to access additional publications. ##Evidence Data START## Transcript over combination: X82200.2, X866728.1, F500002221, ENAacc intron u...

single sampl

**Transcript (**

**Position:**

**Coding Reg**

**Position:**

**Known Genes**

[MN1 \(uc003adj.3\) at chr22:27748277-27801498](#) - Homo sapiens meningioma (disrupted in balanced translocation) 1 (MN1), mRNA. (from RefSeq NM\_002430)

[MN1 \(uc062csm.1\) at chr22:27750064-27796896](#) - The sequence shown here is derived from an Ensembl automatic analysis pipeline and should be considered as preliminary data. (from UniProt H7C105)

[MN1 \(uc010gvg.6\) at chr22:27750678-27791883](#) - meningioma (disrupted in balanced translocation) 1 (from HGNC MN1)

**Page Index**

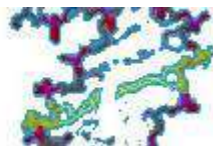
Gene Allele

|                |                   |             |                   |         |
|----------------|-------------------|-------------|-------------------|---------|
| GO Annotations | mRNA Descriptions | Other Names | Model Information | Methods |
|----------------|-------------------|-------------|-------------------|---------|

Data last updated: 2013-06-14

# Базы данных, содержащие информацию о полиморфизмах в генах.

**dbSNP**  
Short Genetic Variations



 NCBI

**SNPedia**



**GWAS Catalog**

- dbSNP
- ClinVar
- SNPedia
- GWAS
- И другие

# dbSNP

- Доступна через интернет по адресу <https://www.ncbi.nlm.nih.gov/SNP/>
- Содержит информацию по полиморфизмам 8 организмов. Для человека представлена информация по 154 206 854 полиморфизмам.
- Идентификаторы начинаются с двух букв rs далее число. Например, полиморфизм в гене человека YY1 имеет идентификатор rs61992955.
- Содержит информацию по
  - ✓ Частотам встречаемости полиморфизма в разных популяциях
  - ✓ Позиции в геноме
  - ✓ Фланкирующим последовательностям
  - ✓ Эффекту и др.

## Информационная карточка полиморфизма.

The screenshot displays the dbSNP website interface. At the top, the NCBI logo and 'dbSNP Short Genetic Variations' are visible. A search bar is present with 'dbSNP' selected. The main content area shows a 'Reference SNP (refSNP) Cluster Report: rs61992955'. The report includes details such as 'Organism: human (Homo sapiens)', 'Molecule Type: Genomic', and 'Created/Updated in build: 129/149'. A table of 'Allele' information is shown, including 'Variation Class: SNV: single nucleotide', 'RefSNP Alleles: C/G (FWD)', and 'Allele Origin: C'. Below this, there are sections for 'SNP Details' with tabs for GeneView, Map, Submission, Fasta, Resource, Diversity, and Validation. An 'Integrated Maps' section provides a table of assembly and annotation data. The 'GeneView' section shows the gene 'YY1' and provides options for clinical source filtering. At the bottom, there are tables for 'Primary Assembly Mapping' and 'RefSeqGene Mapping'.

| Assembly   | Annotation Release | Chr | Chr Pos                   | Contig                       |
|------------|--------------------|-----|---------------------------|------------------------------|
| GRCh38.p7  | 108                | 14  | <a href="#">100274488</a> | <a href="#">NT_026437.13</a> |
| GRCh37.p13 | 105                | 14  | <a href="#">100740825</a> | <a href="#">NT_026437.12</a> |

| RefSeqGene                  | Gene (ID)                  | SNP to RefS |
|-----------------------------|----------------------------|-------------|
| <a href="#">NG_046908.1</a> | <a href="#">YY1 (7528)</a> | Fwd         |

# ClinVar

- Доступна через интернет по адресу <https://www.ncbi.nlm.nih.gov/clinvar/>
- Содержит информацию по полиморфизмам человека и их связи с заболеваниями.
- Идентификаторы – числа, но также используются более разветнутые названия. Например, полиморфизм в гене человека YY1 имеет идентификатор 91950, а также NM\_003403.4(YY1):c.1115C>G (p.Thr372Arg).
- Содержит информацию по
  - ✓ Клинической значимости полиморфизма
  - ✓ Типу полиморфизма
  - ✓ Позиции в геноме и в белке

## Информационная карточка полиморфизма.

The screenshot displays the ClinVar website interface. At the top, there is a search bar and navigation menu. The main content area shows a variant card for NM\_003403.4(YY1):c.1115C>G (p.Thr372Arg). The card includes a variation ID of 91950 and a review status of (0/4) no assertion criteria provided. The clinical significance is listed as 'Uncertain significance'. The interpretation section is currently empty. The allele(s) section shows the variant NM\_003403.4(YY1):c.1115C>G (p.Thr372Arg) with an allele ID of 97428. The variant type is a single nucleotide variant, and the cytogenetic location is 14q32.2. The genomic location is specified as Chr14: 100277470 (on Assembly GRCh38) and Chr14: 100743807 (on Assembly GRCh37). The protein change is T372R. The HGVS notation is provided as NG\_046908.1:g.43706C>G, NM\_003403.4:c.1115C>G, and NP\_003394.1:p.Thr372Arg. Links to UniProtKB (P25490#VAR\_074172) and dbSNP (386834266) are also shown. The NCBI 1000 Genomes Browser link is rs386834266. The molecular consequence is identified as a missense variant. On the right side, there is a section for '1 Affected gene' (YY1 transcription factor (YY1)), a 'Variant frequency in dbGaP' table, and 'Browser views' including RefSeqGene, Variation Viewer, and UCSC. The 'Called variants' table shows 'no data' for called variants and '0 of 4090' for potential variants.

| Sample count | Called variants | Potential variants |
|--------------|-----------------|--------------------|
|              | no data         | 0 of 4090          |



# SNPedia

- Доступна через интернет по адресу <https://www.snpedia.com/>
- Организована в форме Википедии. Любой желающий может вносить и изменять информацию по полиморфизмам человека и их связи с заболеваниями.
- Используются идентификаторы из базы данных dbSNP - начинаются с двух букв rs далее число. Например, полиморфизм в гене человека FCER1A имеет идентификатор rs2494262.
- Содержит информацию по
  - ✓ Клинической значимости полиморфизма
  - ✓ Позиции в геноме
  - ✓ Частотам встречаемости полиморфизма в разных популяциях
  - ✓ Публикациям и др.

## Информационная карточка полиморфизма.

rs2494262

[PMID 19685047] FcepsilonR1alpha gene -18483A>C polymorphism affects transcriptional activity through YY1 binding

[PMID 18846228] Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus.

[PMID 23525950] Single-nucleotide polymorphisms of allergy-related genes and risk of adult glioma

[PMID 25412950] Association of FCER1A genetic polymorphisms

|                     |           |
|---------------------|-----------|
| Orientation         | plus      |
| Stabilized          | plus      |
| Make rs2494262(A:A) |           |
| Make rs2494262(A:C) |           |
| Make rs2494262(C:C) |           |
| Reference           | GRCh38    |
|                     | 38.1/141  |
| Chromosome          | 1         |
| Position            | 159283882 |
| Gene                | FCER1A    |
| GWAS Ctlg           | rs2494262 |
| GMAF                | 0.4601    |
| Max                 |           |
| Magnitude           |           |

? (A:A) (A:C) (C:C) 28

| Population | (A:A) | (A:C) | (C:C) |
|------------|-------|-------|-------|
| CEU        | ~35   | ~45   | ~20   |
| HCB        | ~45   | ~35   | ~20   |
| JPT        | ~55   | ~35   | ~10   |
| YRI        | ~10   | ~40   | ~50   |
| ASW        | ~25   | ~45   | ~30   |
| CHB        | ~45   | ~35   | ~20   |
| CHD        | ~45   | ~35   | ~20   |
| GIH        | ~35   | ~45   | ~20   |
| LWK        | ~10   | ~40   | ~50   |
| MEX        | ~10   | ~40   | ~50   |
| MKK        | ~35   | ~45   | ~20   |
| TSI        | ~35   | ~45   | ~20   |
| AVG        | ~25   | ~45   | ~30   |

# GWAS Catalog

- Доступна через интернет по адресу <https://www.ebi.ac.uk/gwas/>
- Содержит информацию по связи полиморфизмов человека и с заболеваниями и фенотипами.
- Используются идентификаторы из базы данных dbSNP - начинаются с двух букв rs далее число. Например, полиморфизм в гене человека YY1 имеет идентификатор rs2766692.
- Содержит информацию по
  - ✓ Ассоциированным заболеваниям и фенотипам
  - ✓ Позиции в геноме
  - ✓ Частотам встречаемости полиморфизма в разных популяциях
  - ✓ Публикациям и др.

## Информационная карточка полиморфизма.

The screenshot shows the GWAS Catalog search results for the SNP rs2766692. The page includes a search bar with the query 'rs2766692' and a search button. Below the search bar, there are navigation links for 'Home', 'Search', 'Diagram', 'Download', 'Documentation', and 'About'. The main content area displays the GWAS Catalog logo and the text 'The NHGRI-EBI Catalog of published genome-wide association studies'. A search bar contains the query 'rs2766692' and a search button. Below the search bar, there are examples of search results: 'breast cancer, rs7329174, Yang, 2q37.1, HBS1L'. The search results are displayed in a table format. The table has columns for 'Author', 'Date', 'Journal', 'Title', 'Reported trait', and 'Association count'. The first result is for Kang SJ (PMID: 22554406) from Genes Brain Behav, published in 2012-05-03. The title is 'Family-based genome-wide association study of frontal  $\theta$  oscillations identifies potassium channel gene KCNJ6'. The reported trait is 'Electroencephalographic traits in alcoholism' and the association count is 8. The table also includes a 'Refine search results' sidebar on the left with options to show results for 'Studies' (1), 'Associations' (1), and 'Catalog traits' (9). The sidebar also includes a 'Filter results by' section with a p-value filter set to  $\leq 5 \times 10^{-8}$  and an Odds ratio filter set to 'from' and 'to'. The main content area also includes a 'Download association results' button and an 'Expand all studies' button. The bottom of the page shows a list of columns for the search results: 'Author', 'Date', 'Journal', 'Title', 'Reported trait', 'Association count', 'RAF', 'p-value', 'OR', 'Beta', 'CI', 'Region', 'Functional', and 'Reported gene(s)'. The 'Reported gene(s)' column is currently empty.

GWAS / Search / rs2766692

Refine search results

Show results for

- Studies 1
- Associations 1
- Catalog traits 9

Filter results by

p-value  $\leq 5 \times 10^{-8}$

Odds ratio from to

Beta coefficient

Search results for rs2766692

Download association results

Expand all studies

Studies

| Author                   | Date       | Journal           | Title  | Reported trait                               | Association count |
|--------------------------|------------|-------------------|--|--|-------------------|
| Kang SJ (PMID: 22554406) | 2012-05-03 | Genes Brain Behav | Family-based genome-wide association study of frontal $\theta$ oscillations identifies potassium channel gene KCNJ6. | Electroencephalographic traits in alcoholism | 8                 |

Associations

RAF p-value OR Beta CI Region Functional Reported gene(s)

В базе данных dbSNP для гена AGBL4 присутствует информация по 221 431 SNP, а для гена GALE только по 2 843 SNP. Число 221 431 в ~78 раз больше числа 2 843.

Значит ли это что ген AGBL4 в большей степени подвержен мутагенезу?

### Human Gene AGBL4 (uc001cru.2) Description and Page Index

**Description:** Homo sapiens ATP/GTP binding protein-like 4 (AGBL4), mRNA.  
**Transcript (Including UTRs):**  
**Position:** chr1:48,998,527-50,489,626 **Size:** 1,491,100 **Total Exon Count:** 14 **Strand:** -  
**Coding Region:**  
**Position:** chr1:48,999,845-50,489,468 **Size:** 1,489,624 **Coding Exon Count:** 14

SNP

Display Settings: Summary, 20 per page, Sorted by SNP\_ID

Send to:

Search results

Items: 1 to 20 of 221431

1. rs17378497 has merged into rs355206 [Homo sapiens]
1. ATAAAACTGCTTTTCAGTTAAAGG[A/C/T]AAACCTGCTTTTGACCCCTTTCCAAA  
Chromosome: 1:49206502  
Gene: AGBL4 (GeneView)  
Functional Consequence: intron variant  
Validated: by 1000G, by 2hit 2allele, by cluster, by frequency  
Global MAF: A=0.4091/2049  
HGVS: NC\_000001.10:g.49672174C>A, NC\_000001.11:g.49206502C>A, NM\_001323573.1:c.413+39268G>T, NM\_001323574.1:c.413+39268G>T, NM\_001323575.1:c.377+39268G>T, NM\_032785.3:c.377+39268G>T, NR\_136623.1:n.410+39268G>T, XM\_005271284.1:c.413+39268G>T, XM\_011542308.2:c.413+39268G>T, XM\_011542310.2:c.413+39268G>T, XM\_017002595.1:c.377+39268G>T, XM\_017002596.1:c.377+39268G>T, XM\_017002597.1:c.377+39268G>T, XM\_017002598.1:c.377+39268G>T
2. rs17379875 has merged into rs4926831 [Homo sapiens]
2. TGCTAAGCTCTGTGTTGAGCCTCA[C/T]GATTTGTCAAATGATGAGATATGG  
Chromosome: 1:49824429  
Gene: AGBL4 (GeneView)  
Functional Consequence: intron variant  
Validated: by 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap  
Global MAF: T=0.2883/1444  
HGVS: NC\_000001.10:g.50290101C>T, NC\_000001.11:g.49824429C>T, NM\_001323573.1:c.157+26967G>A, NM\_001323574.1:c.157+26967G>A, NM\_001323575.1:c.157+26967G>A, NM\_032785.3:c.157+26967G>A, NR\_136623.1:n.315+26967G>A, XM\_005271284.1:c.157+26967G>A, XM\_011542308.2:c.157+26967G>A, XM\_011542310.2:c.157+26967G>A, XM\_017002595.1:c.157+26967G>A, XM\_017002596.1:c.157+26967G>A, XM\_017002597.1:c.157+26967G>A, XM\_017002598.1:c.157+26967G>A, XM\_017002599.1:c.157+26967G>A

Find related data

Database:

Search details

AGBL4[All Fields]

Recent activity

- AGBL4 (221431)
- Titin (3871)
- Titin TTN (1)
- dystrophin DMD (6)
- dystrophin (7)
- 

### Human Gene GALE (uc001bhx.1) Description and Page Index

**Description:** UDP-galactose-4-epimerase  
**Transcript (Including UTRs):**  
**Position:** chr1:23,994,676-23,999,881 **Size:** 5,206 **Total Exon Count:** 12 **Strand:** -  
**Coding Region:**  
**Position:** chr1:23,995,026-23,998,084 **Size:** 3,059 **Coding Exon Count:** 10

SNP

Display Settings: Summary, 20 per page, Sorted by SNP\_ID

Send to:

Search results

Items: 1 to 20 of 2843

- rs28940885 [Homo sapiens]
1. CCCAGCCTGGCCCAAGGAGGACTGG[A/G]GTGGACAGCAGCCTTAGGGCTGGAC  
Chromosome: 1:23796183  
Gene: GALE (GeneView)  
Functional Consequence: missense  
Allele Origin: G(germline)/A(germline)  
Clinical significance: Pathogenic  
Validated: by 1000G, by cluster, by frequency  
Global MAF: T=0.0020/10  
HGVS: NC\_000001.10:g.24122673C>T, NC\_000001.11:g.23796183C>T, NG\_007068.1:g.9622G>A, NM\_000403.3:c.956G>A, NM\_001008216.1:c.956G>A, NM\_001127621.1:c.956G>A, NP\_000394.2:p.Gly319Glu, NP\_001008217.1:p.Gly319Glu, NP\_001121093.1:p.Gly319Glu, XM\_005245833.1:c.1066G>A, XM\_005245834.1:c.1066G>A, XM\_005245835.1:c.1066G>A, XM\_005245836.1:c.874G>A, XM\_005245837.1:c.764G>A, XP\_005245890.1:p.Gly356Ser, XP\_005245891.1:p.Gly356Ser, XP\_005245892.1:p.Gly356Ser, XP\_005245893.1:p.Gly292Ser, XP\_005245894.1:p.Gly255Glu
- rs28940884 [Homo sapiens]
2. AAGGCCACATTCGACCTTAAGGA[A/G]GCTGAAAGAACAGTGTGGCTGCCGG  
Chromosome: 1:23796722  
Gene: GALE (GeneView)  
Functional Consequence: missense  
Allele Origin: G(germline)/A(germline)  
Clinical significance: Pathogenic  
Validated: by 1000G, by cluster, by frequency  
Global MAF: C=0.0082/41  
HGVS: NC\_000001.10:g.24123212T>C, NC\_000001.11:g.23796722T>C, NG\_007068.1:g.9083A>G, NM\_000403.3:c.770A>G, NM\_001008216.1:c.770A>G, NM\_001127621.1:c.770A>G, NP\_000394.2:p.Val297Asn, NP\_001008217.1:n.Val257Asn

Find related data

Database:

Search details

GALE[All Fields]

Recent activity

- GALE (2843)
- AGBL4 (221431)
- Titin (3871)
- Titin TTN (1)
- dystrophin DMD (6)
-

В базе данных dbSNP для гена **AGBL4** присутствует информация по 221 431 SNP, а для гена **GALE** только по 2 843 SNP. Число 221 431 в ~78 раз больше числа 2 843.

Значит ли это что ген **AGBL4** в большей степени подвержен мутагенезу?

Длина гена **AGBL4** составляет 1 491 100 п.н., а гена **GALE** – 5 206 п.н., т.е. ген **AGBL4** в 286 раз длиннее гена **GALE**. Таким образом, в гене **AGBL4** на 100 п.н. приходится 14,85 SNP, а в гене **GALE** – 54,61, т.е. число известных мутаций на п.н. в гене **GALE** значительно превосходит это же число для гена **AGBL4**.

### Human Gene **AGBL4** (uc001cru.2) Description and Page Index

**Description:** Homo sapiens ATP/GTP binding protein-like 4 (AGBL4), mRNA.  
**Transcript (Including UTRs):**  
**Position:** chr1:48,998,527-50,489,624 **Size: 1,491,100** **Total Exon Count:** 14 **Strand:** -  
**Coding Region:**  
**Position:** chr1:48,999,845-50,489,468 **Size: 1,489,624** **Coding Exon Count:** 14

### Human Gene **GALE** (uc001bhx.1) Description and Page Index

**Description:** UDP-galactose-4-epimerase  
**Transcript (Including UTRs):**  
**Position:** chr1:23,994,676-23,999,882 **Size: 5,206** **Total Exon Count:** 12 **Strand:** -  
**Coding Region:**  
**Position:** chr1:23,995,026-23,998,084 **Size: 3,059** **Coding Exon Count:** 10

SNP  Search  
Create alert Advanced Help

SNP  Search  
Create alert Advanced Help

Display Settings: Summary, 20 per page, Sorted by SNP\_ID  
Send to: Filters: Manage Filters

Search results: Items: 1 to 20 of 221431

1. rs17378497 has merged into rs355206 [Homo sapiens]  
1. ATAAAACTGCTTTTCAGTTAAGG[A/C]AAACCTGCTTTTGACCTTTCCAAA  
Chromosome: 1:49206502  
Gene: **AGBL4** (GeneView)  
Functional Consequence: intron variant  
Validated: by 1000G, by 2hit 2allele, by cluster, by frequency  
Global MAF: A=0.4091/2049  
HGVS: NC\_000001.10:g.49672174C>A, NC\_000001.11:g.49206502C>A, NM\_001323573.1:c.413+39268G>T, NM\_001323574.1:c.413+39268G>T, NM\_001323575.1:c.377+39268G>T, NM\_032785.3:c.377+39268G>T, NR\_136623.1:n.410+39268G>T, XM\_005271284.1:c.413+39268G>T, XM\_011542308.2:c.413+39268G>T, XM\_011542310.2:c.413+39268G>T, XM\_017002595.1:c.377+39268G>T, XM\_017002596.1:c.377+39268G>T, XM\_017002597.1:c.377+39268G>T, XM\_017002598.1:c.377+39268G>T

2. rs17379875 has merged into rs4926831 [Homo sapiens]  
2. TGCTAAGCTCTGTGTTGAGCCTCA[C/T]GATTTGTCAAATGATGAGATATGG  
Chromosome: 1:49824429  
Gene: **AGBL4** (GeneView)  
Functional Consequence: intron variant  
Validated: by 1000G, by 2hit 2allele, by frequency, by hapmap  
Global MAF: T=0.2883/1444  
HGVS: NC\_000001.10:g.50290101C>T, NC\_000001.11:g.49824429C>T, NM\_001323573.1:c.157+26967G>A, NM\_001323574.1:c.157+26967G>A, NM\_001323575.1:c.157+26967G>A, NM\_032785.3:c.157+26967G>A, NR\_136623.1:n.315+26967G>A, XM\_005271284.1:c.157+26967G>A, XM\_011542308.2:c.157+26967G>A, XM\_011542310.2:c.157+26967G>A, XM\_017002595.1:c.157+26967G>A, XM\_017002596.1:c.157+26967G>A, XM\_017002597.1:c.157+26967G>A, XM\_017002598.1:c.157+26967G>A

Find related data  
Database: Select  
Find items

Search details  
AGBL4[All Fields]  
Search See more...

Recent activity  
AGBL4 (221431) SNP  
Titin (3871) SNP  
Titin TTN (1) SNP  
dystrophin DMD (6) SNP  
dystrophin (7) SNP  
See more...

Display Settings: Summary, 20 per page, Sorted by SNP\_ID  
Send to: Filters: Manage Filters

Search results: Items: 1 to 20 of 2843

rs28940885 [Homo sapiens]  
1. CCCAGCTGGCCCAAGGAGGCTGG[A/G]GTGGACAGCAGCCTTAGGGCTGGAC  
Chromosome: 1:23796183  
Gene: **GALE** (GeneView)  
Functional Consequence: missense  
Allele Origin: G(germline)/A(germline)  
Clinical significance: Pathogenic  
Validated: by 1000G, by cluster, by frequency  
Global MAF: T=0.0020/10  
HGVS: NC\_000001.10:g.24122673C>T, NC\_000001.11:g.23796183C>T, NG\_007068.1:g.9622G>A, NM\_000403.3:c.956G>A, NM\_001008216.1:c.956G>A, NM\_001127621.1:c.956G>A, NP\_000394.2:p.Gly319Glu, NP\_001008217.1:p.Gly319Glu, NP\_001121093.1:p.Gly319Glu, XM\_005245833.1:c.1066G>A, XM\_005245834.1:c.1066G>A, XM\_005245835.1:c.1066G>A, XM\_005245836.1:c.874G>A, XM\_005245837.1:c.764G>A, XP\_005245890.1:p.Gly356Ser, XP\_005245891.1:p.Gly356Ser, XP\_005245892.1:p.Gly356Ser, XP\_005245893.1:p.Gly292Ser, XP\_005245894.1:p.Gly255Glu

rs28940884 [Homo sapiens]  
2. AAGGGCCACATTCGACCTTAAAGGA[A/G]GCTGAAAGAACAGTGTGGCTGCCGG  
Chromosome: 1:23796722  
Gene: **GALE** (GeneView)  
Functional Consequence: missense  
Allele Origin: G(germline)/A(germline)  
Clinical significance: Pathogenic  
Validated: by 1000G, by cluster, by frequency  
Global MAF: C=0.0082/41  
HGVS: NC\_000001.10:g.24123212T>C, NC\_000001.11:g.23796722T>C, NG\_007068.1:g.9683A>G, NM\_000403.3:c.770A>G, NM\_001008216.1:c.770A>G, NM\_001127621.1:c.770A>G, NP\_000394.2:p.Val297Asn, NP\_001008217.1:n.Val257Asn

Find related data  
Database: Select  
Find items

Search details  
GALE[All Fields]  
Search See more...

Recent activity  
GALE (2843) SNP  
AGBL4 (221431) SNP  
Titin (3871) SNP  
Titin TTN (1) SNP  
dystrophin DMD (6) SNP  
See more...

# Базы данных, содержащие информацию по белкам.



- UniProt
- NCBI Protein
- PDB
- И другие

# Uniprot (Universal Protein Resource)

- Доступна через интернет по адресу <http://www.uniprot.org/>
- Содержит информацию по структуре и функциям белков 457010 бактерий, 168308 вирусов, 12163 архей, 894013 эукариот.
- База содержит два раздела: 1) Swiss-Prot, включающий карточки белков, проаннотированные вручную экспертами; 2) TrEMBL, включающий карточки белков, сформированные автоматическими средствами.
- Идентификаторы двух типов 1) шесть знаков: цифры и заглавные буквы 2) короткое название белка\_организм. Например, белок YY1 человека P25490 и TYY1\_HUMAN, мыши - Q00899 и TYY1\_MOUSE.
- Содержит информацию по
  - ✓ синонимам и функциям
  - ✓ Локализации в клетке
  - ✓ Пост-трансляционным модификациям
  - ✓ Экспрессии
  - ✓ Взаимодействиям белка
  - ✓ Последовательности и др.

## Информационная карточка белка.

UniProtKB - P25490 (TYY1\_HUMAN)

Protein | **Transcriptional repressor protein YY1**

Gene | **YY1**

Organism | *Homo sapiens (Human)*

Status | Reviewed - Annotation score: ●●●●●● - Experimental evidence at protein level<sup>1</sup>

### Function<sup>1</sup>

Multifunctional transcription factor that exhibits positive and negative control on a large number of cellular and viral genes by binding to sites overlapping the transcription start site. Binds to the consensus sequence 5'-CCGCCATNTT-3'; some genes have been shown to contain a longer binding motif allowing enhanced binding; the initial CG dinucleotide can be methylated greatly reducing the binding affinity. The effect on transcription regulation is depending upon the context in which it binds and diverse mechanisms of action include direct activation or repression, indirect activation or repression via cofactor recruitment, or activation or repression by disruption of binding sites or conformational DNA changes. Its activity is regulated by transcription factors and cytoplasmic proteins that have been shown to abrogate or completely inhibit YY1-mediated activation or repression. For example, it acts as a repressor in absence of adenovirus E1A protein but as an activator in its presence. Acts synergistically with the SMAD1 and SMAD4 in bone morphogenetic protein (BMP)-mediated cardiac-specific gene expression (PubMed:15329343). Binds to SMAD binding elements (SBEs) (5'-GTCT/AGAC-3') within BMP response element (BMPRE) of cardiac activating regions. May play an important role in development and differentiation. Proposed to recruit the PRC2/EED-EZH2 complex to target genes that are transcriptionally repressed. Involved in DNA repair. In vitro, binds to DNA recombination intermediate structures (Holliday junctions). [3 Publications](#)

Proposed core component of the chromatin remodeling INO80 complex which is involved in transcriptional regulation, DNA replication and probably DNA repair; proposed to target the INO80 complex to YY1-responsive elements. [2 Publications](#)

Sites

None

- Function
- Names & Taxonomy
- Subcell. location
- Pathol./Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Cross-references
- Entry information

# NCBI Protein

- Доступна через интернет по адресу <https://www.ncbi.nlm.nih.gov/protein/>
- Содержит информацию по аминокислотным последовательностям белков различных организмов.
- Используются идентификаторы различных баз данных, таких как GenBank и RefSeq. Например, белок YY1 человека имеет идентификатор AAH65366.1
- Содержит информацию по
  - ✓ Организмам, таксономическому положению организма
  - ✓ Ссылки на публикации
  - ✓ Разметке функциональных сайтов белка
  - ✓ Аминокислотной последовательности и др.

## Информационная карточка белка.

The screenshot shows the NCBI Protein database entry for YY1 transcription factor [Homo sapiens]. The page includes a search bar, navigation options, and a detailed description of the protein. The amino acid sequence is displayed at the bottom.

**GenPept** Send to: ▾ Change region shown ▾ Customize view ▾

**YY1 transcription factor [Homo sapiens]**  
GenBank: AAH65366.1  
[Identical Proteins](#) [FASTA](#) [Graphics](#)

[Go to: ▾](#)

**LOCUS** AAH65366 414 aa linear PRI 15-JUL-2006  
**DEFINITION** YY1 transcription factor [Homo sapiens].  
**ACCESSION** AAH65366  
**VERSION** AAH65366.1  
**DBSOURCE** accession [BC065366.1](#)  
**KEYWORDS** MGC.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** [Homo sapiens](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (residues 1 to 414)  
**AUTHORS** Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,B., Lippman,H., Moore,T., May,S.T., Wang,J., Hojima,F., et al.

**Region**  
/db\_xref="CDD:296206"  
385..407  
/region\_name="C2H2 Zn finger"  
/note="C2H2 Zn finger [structural motif]"  
/db\_xref="CDD:275368"  
order(385,390,403,407)  
/site\_type="other"  
/note="Zn binding site [ion binding]"  
/db\_xref="CDD:275368"  
1..414  
/gene="YY1"  
/gene\_synonym="DELTA"  
/gene\_synonym="NF-E1"  
/gene\_synonym="UCRBP"  
/gene\_synonym="YIN-YANG-1"  
/coded\_by="BC065366.1:46..1290"  
/db\_xref="GeneID:7528"  
/db\_xref="HGNC:HGNC:12856"  
/db\_xref="MIM:600013"

**ORIGIN**  
1 masgdtlyia tdgsempaei velheievet ipvetiettv vgeeeeeedd dedggggdhg  
61 gggghghag hhhhhhhhh ppmaalqlplv tddptqvhhh qeivilvqtr evvggdsdsg  
121 laedgfdedq ilipvpapag gddyieqtl vtvaagksg gggsssggg rvkkgggkks  
181 gkksylsgga gaaggggpad gnkkeqkqv qiktlegefs vtmwssdekk didhetvvee  
241 qiigenspdp yseymtgkkl ppggipgidl sdpkqlaefa rmkprkiked daprtiacph  
301 kgctkmfrdn samrkhhlth gprvhvcaec gkafvesskl krhqlvhtge kpfqctfegc  
361 gkrfsldfnl rthvrihtgd rpyvcpfdgc nkkfaqstnl kshilthaka knnq

**Protein 3D Structure**  
  
Crystal Structure Of Mbd1 Yy1 Complex  
PDB: 4C5I  
Source: Homo sapiens

**LinkOut to external resources**  
A selection of literature about the proteins [\[GoPubMed Proteins\]](#)

**MODBASE, Database of Comparative Pi** [\[MODBASE, Database of Comparat..\]](#)

**Transcript/Protein Information** [\[PANTHER Classification System\]](#)

**Protein Ontology Consortium** [\[Protein Ontology Consortium\]](#)

**reagents** [\[ExactAntigen/Labome\]](#)

**reagent reviews** [\[ExactAntigen/Labome\]](#)

**Recent activity**  
[Turn Off](#) [Clear](#)

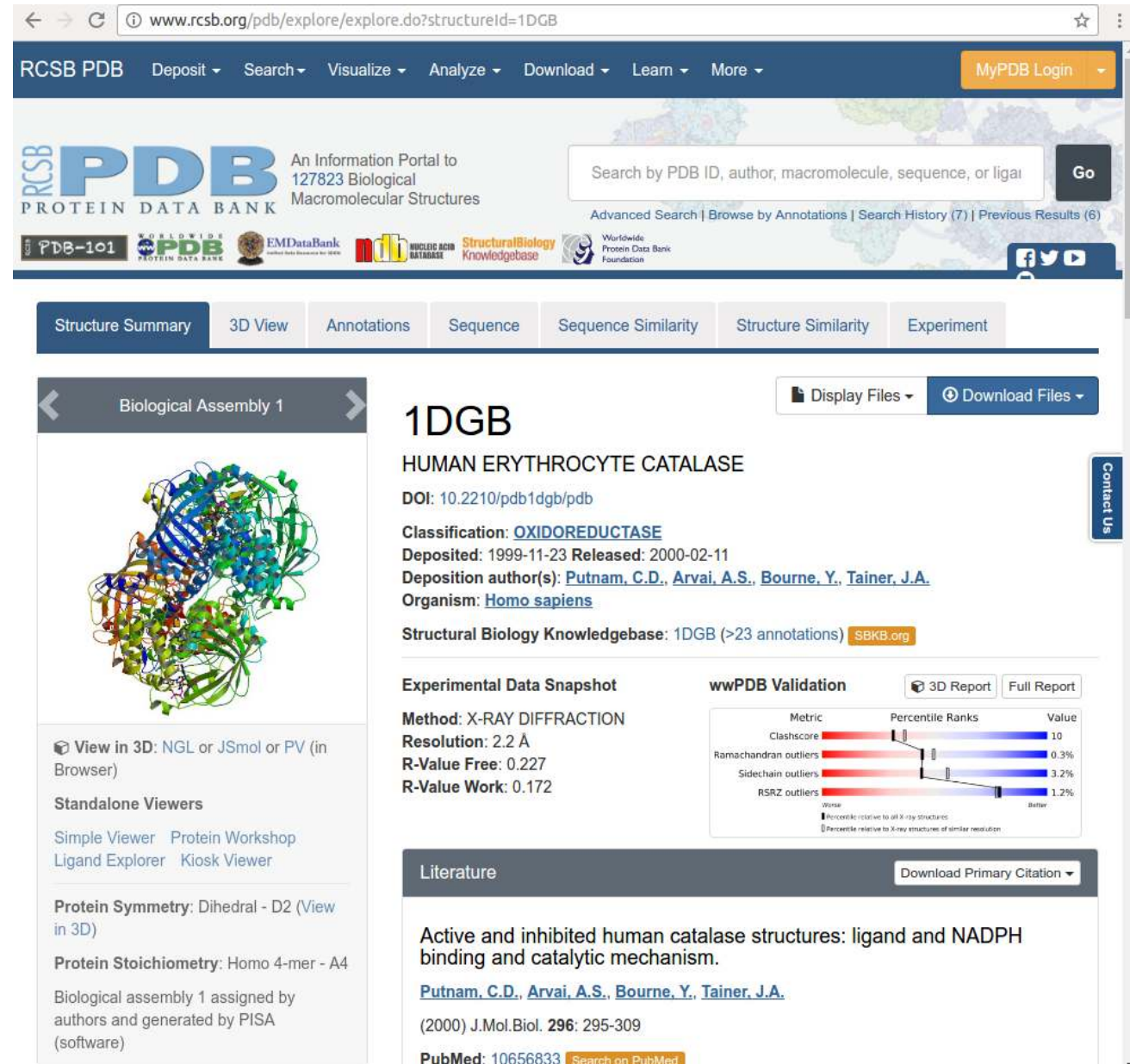
YY1 transcription factor [Homo sapiens] Protein

(yy1) AND "Homo sapiens"[porgn] Protein

# PDB - Protein Data Bank

- Доступна через интернет по адресу <http://www.rcsb.org/pdb/home/home.do>
- Содержит информацию по 3D структуре белков и белковых комплексов различных организмов.
- Идентификаторы четыре символа: цифры и заглавные буквы. Например, белок HUMAN ERYTHROCYTE CATALASE имеет идентификатор 1DGB.
- Содержит информацию по
  - ✓ Функции белка
  - ✓ Ссылки на публикации
  - ✓ Лигандам
  - ✓ 3D структуре
  - ✓ Аминокислотной последовательности
  - ✓ Вторичной структуре белка и др.

## Информационная карточка белка.



www.rcsb.org/pdb/explore/explore.do?structureid=1DGB

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 127823 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligand Go

Advanced Search | Browse by Annotations | Search History (7) | Previous Results (6)

PDB-101 WORLDWIDE PDB PROTEIN DATA BANK EMDatabank NUCLEIC ACID DATABASE StructuralBiology Knowledgebase Worldwide Protein Data Bank Foundation

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment

Biological Assembly 1

1DGB

HUMAN ERYTHROCYTE CATALASE

DOI: 10.2210/pdb1dgb/pdb

Classification: **OXIDOREDUCTASE**

Deposited: 1999-11-23 Released: 2000-02-11

Deposition author(s): [Putnam, C.D.](#), [Arvai, A.S.](#), [Bourne, Y.](#), [Tainer, J.A.](#)

Organism: [Homo sapiens](#)

Structural Biology Knowledgebase: 1DGB (>23 annotations) [SBKB.org](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.2 Å

R-Value Free: 0.227

R-Value Work: 0.172

wwPDB Validation

| Metric                | Percentile Ranks | Value |
|-----------------------|------------------|-------|
| Clashscore            |                  | 10    |
| Ramachandran outliers |                  | 0.3%  |
| Sidechain outliers    |                  | 3.2%  |
| RSRZ outliers         |                  | 1.2%  |

Literature

Download Primary Citation

Active and inhibited human catalase structures: ligand and NADPH binding and catalytic mechanism.

[Putnam, C.D.](#), [Arvai, A.S.](#), [Bourne, Y.](#), [Tainer, J.A.](#)

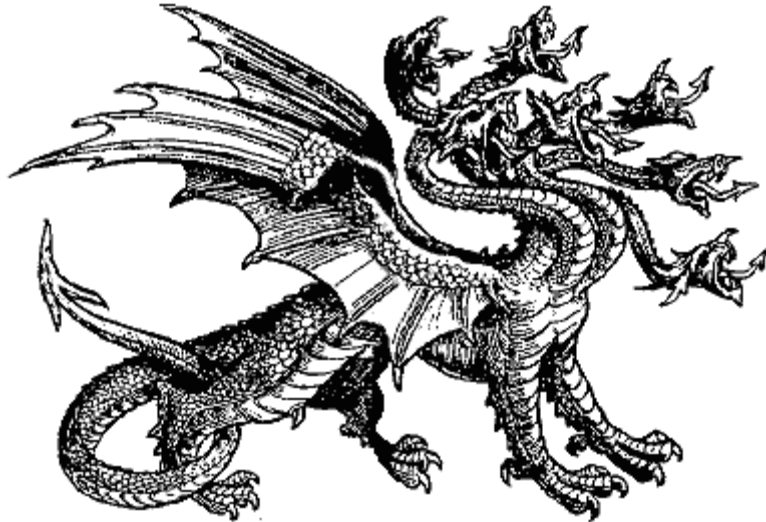
(2000) J.Mol.Biol. **296**: 295-309

PubMed: 10656833 [Search on PubMed](#)



Обычно в базе данных UniProt для каждого белка существует отдельная информационная карточка. Однако, существуют записи, для которых представлено большое число основных равноправных названий белков. Например, в информационной карточке P27958 (POLG\_HCVH) 11 рекомендованных названий белков.

С чем может быть связано такое большое число основных равноправных названий белков, описанных в одной информационной карточке?



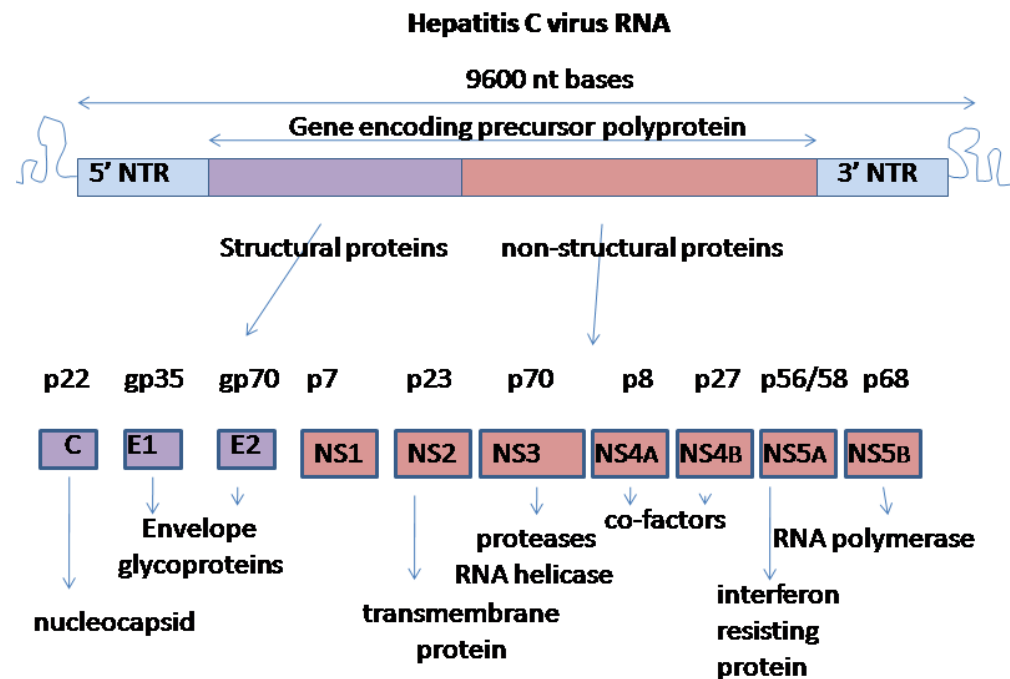
## Информационная карточка UniProtKB - P27958 (POLG\_HCVH)

| Protein names <sup>1</sup> | Recommended name:  |
|----------------------------|--|
|                            | <ul style="list-style-type: none"><li>• <b>Core protein p21</b><br/><i>Alternative name(s):</i><ul style="list-style-type: none"><li>▪ Capsid protein C</li><li>▪ p21</li></ul></li><li>• <b>Core protein p19</b></li><li>• <b>Envelope glycoprotein E1</b><br/><i>Alternative name(s):</i><ul style="list-style-type: none"><li>▪ gp32</li><li>▪ gp35</li></ul></li><li>• <b>Envelope glycoprotein E2</b><br/><i>Alternative name(s):</i><ul style="list-style-type: none"><li>▪ NS1</li><li>▪ gp68</li><li>▪ gp70</li></ul></li><li>• <b>p7</b></li><li>• <b>Protease NS2-3 (EC:3.4.22.-)</b><ul style="list-style-type: none"><li>▪ <i>Short</i> p23<br/><i>name:</i> • <b>Serine protease NS3 (EC:3.4.21.98, EC:3.6.1.15, EC:3.6.4.13)</b></li></ul><i>Alternative name(s):</i><ul style="list-style-type: none"><li>▪ Hepacivirin</li><li>▪ NS3P</li><li>▪ p70</li></ul></li><li>• <b>Non-structural protein 4A</b><ul style="list-style-type: none"><li>▪ <i>Short</i> NS4A<br/><i>name:</i> <i>Alternative name(s):</i></li><li>▪ p8</li></ul></li><li>• <b>Non-structural protein 4B</b><ul style="list-style-type: none"><li>▪ <i>Short</i> NS4B<br/><i>name:</i> <i>Alternative name(s):</i></li><li>▪ p27</li></ul></li><li>• <b>Non-structural protein 5A</b><ul style="list-style-type: none"><li>▪ <i>Short</i> NS5A<br/><i>name:</i> <i>Alternative name(s):</i></li><li>▪ p56</li></ul></li><li>• <b>RNA-directed RNA polymerase (EC:2.7.7.48)</b><br/><i>Alternative name(s):</i><ul style="list-style-type: none"><li>▪ NS5B</li><li>▪ p68</li></ul></li></ul> |
| Organism <sup>1</sup>      | Hepatitis C virus genotype 1a (isolate H) (HCV)  |

Обычно в базе данных UniProt для каждого белка существует отдельная информационная карточка. Однако, существуют записи, для которых представлено большое число основных равноправных названий белков. Например, в информационной карточке P27958 (POLG\_HCVH) 11 рекомендованных названий белков.

С чем может быть связано такое большое число основных равноправных названий белков, описанных в одной информационной карточке?

## Информационная карточка UniProtKB - P27958 (POLG\_HCVH)



Protein names

**Genome polyprotein**  
Cleaved into the following 11 chains:

- Alternative name(s):
  - Capsid protein C
  - p21
- Core protein p19
- Envelope glycoprotein E1
  - Alternative name(s):
    - gp32
    - gp35
- Envelope glycoprotein E2
  - Alternative name(s):
    - NS1
    - gp68
    - gp70
- p7
- Protease NS2-3 (EC:3.4.22.-)
  - Short name: p23
  - name: Serine protease NS3 (EC:3.4.21.98, EC:3.6.1.15, EC:3.6.4.13)
  - Alternative name(s):
    - Hepacivirin
    - NS3P
    - p70
- Non-structural protein 4A
  - Short name: NS4A
  - name: Alternative name(s):
  - p8
- Non-structural protein 4B
  - Short name: NS4B
  - name: Alternative name(s):
  - p27
- Non-structural protein 5A
  - Short name: NS5A
  - name: Alternative name(s):
  - p56
- RNA-directed RNA polymerase (EC:2.7.7.48)
  - Alternative name(s):
    - NS5B
    - p68

Organism<sup>1</sup> Hepatitis C virus genotype 1a (isolate H) (HCV)

# Базы данных, содержащие информацию по микроРНК.



- miRBase
- miRDB
- miRTarBase
- И другие

# miRBase

- Доступна через интернет по адресу <http://www.mirbase.org/>
- Содержит информацию по структуре и последовательности микроРНК 223 организмов.
- Идентификаторы: MI затем семь цифр. Например, микроРНК мыши mmu-mir-302b имеет идентификатор MI0003716.
- Содержит информацию по
  - ✓ Последовательности шпильки
  - ✓ Ссылки на публикации
  - ✓ Расположению в геноме
  - ✓ Зрелым микроРНК и др.

## Информационная карточка микроРНК.

www.mirbase.org/cgi-bin/mirna\_entry.pl?acc=MI0003716


miRBase MANCHESTER 1824

Home Search Browse Help Download Blog Submit mmu-mir-302b Search

### Stem-loop sequence mmu-mir-302b

|                      |   |                         |  |                         |  |
|----------------------|---|-------------------------|--|-------------------------|--|
| Accession            | MI0003716   |                         |  |                         |  |
| Symbol               | <a href="#">MGI:Mir302b</a>   |                         |  |                         |  |
| Description          | Mus musculus miR-302b stem-loop   |                         |  |                         |  |
| Gene family          | MIPF0000071; <a href="#">mir-302</a>  |                         |  |                         |  |
| Stem-loop            | <pre>      guucc  a  uu      au  ucugu  au 5'      cuuc  acu  aacauggga  gcuu  cuc  c                                   3'      gaag  uga  uuguaccuu  ugaa  gag  g       ---u  a  uu      cg  ---u  aa</pre> <p><a href="#">Get sequence</a></p> <p><a href="#">1633</a> reads, <a href="#">84.5</a> reads per million, 35 experiments</p>  |                         |  |                         |  |
| Deep sequencing      |   |                         |  |                         |  |
| Confidence           | Feedback: Do you believe this miRNA is real? <input type="button" value="Yes (+4)"/> <input type="button" value="No (-0)"/> <input type="button" value="Leave comment"/>  |                         |  |                         |  |
| Comments             | Mouse miR-302b was verified experimentally by Mineno et al using MPSS technology [1]. The MPSS protocol used provides 22nt sequences, but the true extents of the mature miRNA are not reliably obtained. The mature sequence shown here represents the most commonly cloned form from large-scale cloning studies [2]. The 5' end of the miRNA may be offset with respect to previous annotations. |                         |  |                         |  |
| Genome context       | <table><tr><td>Coordinates (GRCm38)</td><td>chr3: <a href="#">127545228-127545301</a> [+]</td><td>Overlapping transcripts</td><td>antisense <a href="#">ENSMUST00000029588</a>; Larp7-201; intron 8</td></tr></table>   | Coordinates (GRCm38)    | chr3: <a href="#">127545228-127545301</a> [+]                      | Overlapping transcripts | antisense <a href="#">ENSMUST00000029588</a> ; Larp7-201; intron 8 |
| Coordinates (GRCm38) | chr3: <a href="#">127545228-127545301</a> [+]   | Overlapping transcripts | antisense <a href="#">ENSMUST00000029588</a> ; Larp7-201; intron 8 |                         |  |

- Доступна через интернет по адресу <http://www.mirdb.org/miRDB/index.html>
- Содержит информацию по структуре и мишеням микроРНК человека, мыши, крысы, собаки и курицы.
- Идентификаторы: название микроРНК. Например, hsa-miR-302b-3p.
- ✓ Приставка «mir» отделяется дефисом, вслед за ней следует номер, говорящий о порядке именованя (mir-123 была открыта и названа раньше, чем mir-302).
- ✓ «mir-» обозначает пре-микроРНК, «MIR-» ген, кодирующий микроРНК, а «miR-» — для обозначения зрелой формы.
- ✓ К названию микроРНК с последовательностями, отличающимися на один или два нуклеотида, приписывается строчная буква (miR-123a и miR-123b).
- ✓ Пре-микроРНК, дающие начало на 100 % идентичным микроРНК, но локализованные в разных местах генома, имеют в названии цифру, отделенную дефисом (hsa-mir-194-1 и hsa-mir-194-2).
- ✓ Вид, из которого была выделена микроРНК, обозначается в названии трёхбуквенной приставкой (hsa-miR-123 человека).
- ✓ Когда две зрелые микроРНК образуются из двух различных концов исходной пре-микроРНК, к ним добавляется суффикс -3p или -5p.
- ✓ Когда известен относительный уровень экспрессии для двух микроРНК, имеющих общего предшественника, тогда микроРНК, которая экспрессируется на более низком уровне, чем микроРНК с противоположного конца шпильки, помечают звёздочкой (miR-123 и miR-123\* имеют общую исходную шпилечную пре-микроРНК, но в клетке обнаруживается больше miR-123).




**MicroRNA and Target Gene Description:**

|                         |                       |                          |                             |
|-------------------------|-----------------------|--------------------------|-----------------------------|
| <b>miRNA Name</b>       | hsa-miR-302b-3p       | <b>miRNA Sequence</b>    | UAGUGUCUCCAUGUUUUAGUAG      |
| <b>Previous Name</b>    | hsa-miR-302b          |                          |                             |
| <b>Target Score</b>     | 100                   | <b>Seed Location</b>     | 984, 1137, 1538, 1600, 3494 |
| <b>NCBI Gene ID</b>     | <a href="#">55432</a> | <b>GenBank Accession</b> | <a href="#">NM_018566</a>   |
| <b>Gene Symbol</b>      | YOD1                  | <b>3' UTR Length</b>     | 5171                        |
| <b>Gene Description</b> | YOD1 deubiquitinase   |                          |                             |

**3' UTR Sequence**

```

1 CCTATGCATG AATGAGGGTT GAAGCCTACT ACCTCACACA TCCAGAAGGC TCTGGGTTTT
61 CCAATAAGCT ATGGTAACCC TAAAGAACAA AGGATACAAT GCTTGAACCA TCCTTTTAAAC
121 TTAATAACAC TAAGACACTG AAATTCCTTG TTAAGATTAA AATTAGTGTG CAAGTTTACA
181 GATGTGTGTC TACAGTGGTA AACTGTACAT ACATGCCTCT TTCTGCTGGA GTGACAGAAT
241 AGGTGATCCT TGGACCTAC TGACACTGAC CTGAAGGTTG AGATTGAGTA TTATAAACTA
    
```



**There are 615 predicted targets for hsa-miR-302b-3p in miRDB.**

| Target Detail           | Target Rank | Target Score | miRNA Name      | Gene Symbol            | Gene Description                         |
|-------------------------|-------------|--------------|-----------------|------------------------|--|
| <a href="#">Details</a> | 1           | 100          | hsa-miR-302b-3p | <a href="#">YOD1</a>   | YOD1 deubiquitinase                      |
| <a href="#">Details</a> | 2           | 100          | hsa-miR-302b-3p | <a href="#">OXR1</a>   | oxidation resistance 1                   |
| <a href="#">Details</a> | 3           | 100          | hsa-miR-302b-3p | <a href="#">LATS2</a>  | large tumor suppressor kinase 2          |
| <a href="#">Details</a> | 4           | 100          | hsa-miR-302b-3p | <a href="#">NR2C2</a>  | nuclear receptor subfamily 2, group C, 2 |
| <a href="#">Details</a> | 5           | 100          | hsa-miR-302b-3p | <a href="#">ZNF800</a> | zinc finger protein 800                  |
| <a href="#">Details</a> | 6           | 100          | hsa-miR-302b-3p | <a href="#">CROT</a>   | carnitine O-octanoyltransferase          |
| <a href="#">Details</a> | 7           | 100          | hsa-miR-302b-3p | <a href="#">CYBRD1</a> | cytochrome b reductase 1                 |
| <a href="#">Details</a> | 8           | 99           | hsa-miR-302b-3p | <a href="#">ZNF367</a> | zinc finger protein 367                  |
| <a href="#">Details</a> | 9           | 99           | hsa-miR-302b-3p | <a href="#">REEP3</a>  | receptor accessory protein 3             |
| <a href="#">Details</a> | 10          | 99           | hsa-miR-302b-3p | <a href="#">RSBN1</a>  | round spermatid basic protein 1          |
| <a href="#">Details</a> | 11          | 99           | hsa-miR-302b-3p | <a href="#">MPC1</a>   | mitochondrial pyruvate carrier 1         |
| <a href="#">Details</a> | 12          | 99           | hsa-miR-302b-3p | <a href="#">PPP6C</a>  | protein phosphatase 6, catalytic subun   |

# miRTarBase

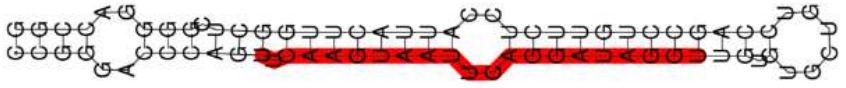

- Доступна через интернет по адресу <http://mirtarbase.mbc.nctu.edu.tw/>
- Содержит информацию по экспериментально подтвержденным взаимодействиям микроРНК-мишень для 18 организмов.
- Идентификаторы: MIRT затем шесть цифр. Например, микроРНК человека hsa-miR-26b-5p имеет идентификатор MIRT029499.
- Содержит информацию по
  - ✓ Мишеням
  - ✓ Последовательности шпильки
  - ✓ Синонимам
  - ✓ Ассоциированным заболеваниям
  - ✓ Ссылки на публикации и др.

## Информационная карточка микроРНК.

**miRTarBase**

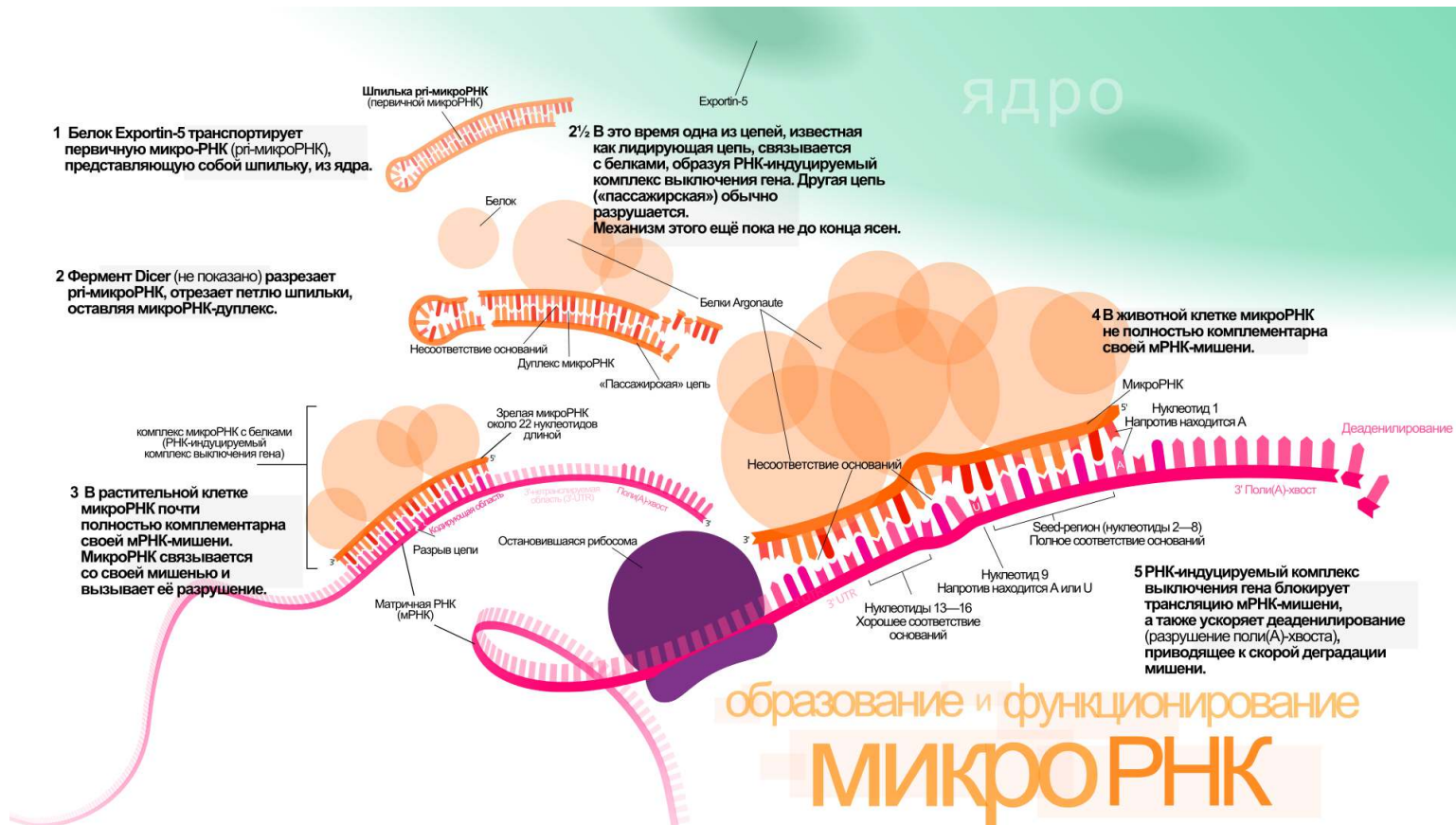
Accession ID: MIRT029499 [miRNA, hsa-miR-26b-5p :: MIR22HG, target gene]

miRNA Target Gene Evidences Expression TCGA Gene Set Enrichments Network ERROR Report

| pre-miRNA Information      |  |
|----------------------------|--|
| pre-miRNA ID               | hsa-mir-26bLinkOut: [miRBase]  |
| Synonyms                   | MIRN26B, hsa-mir-26b, miR-26b, MIR26B  |
| Description                | Homo sapiens miR-26b stem-loop   |
| Comment                    | The mature sequence shown here represents the most commonly cloned form from large-scale cloning |
| 2nd Structure of pre-miRNA |               |
| Disease                    |             |

МикроРНК (англ. microRNA, miRNA) — малые некодирующие молекулы РНК длиной 18—25 нуклеотидов (в среднем 22), принимающие участие в транскрипционной и посттранскрипционной регуляции экспрессии генов путём РНК-интерференции.

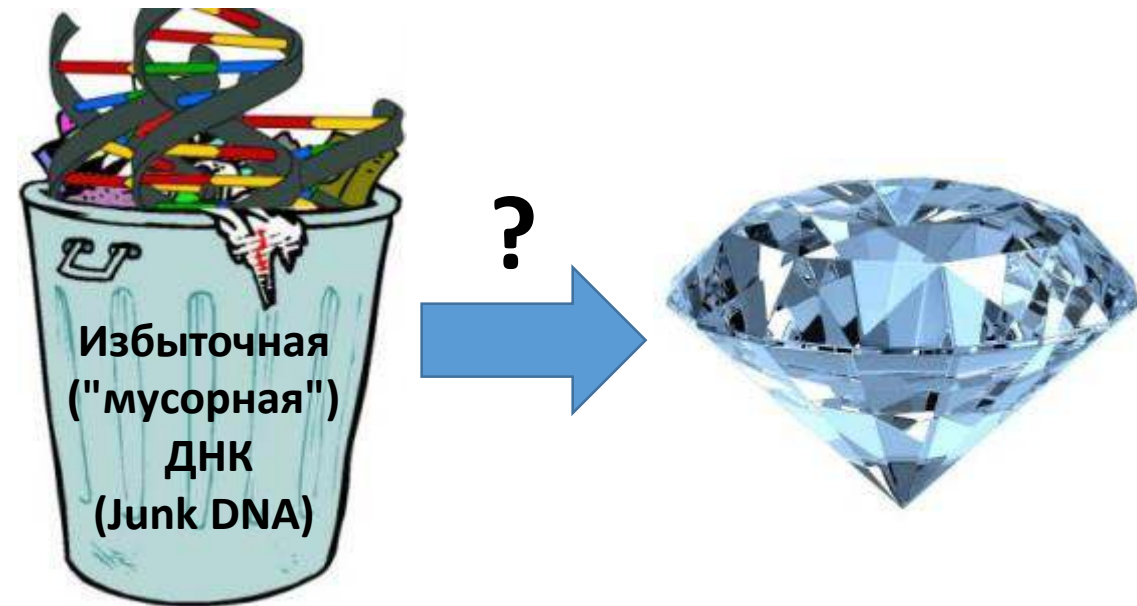
А какие еще классы некодирующих РНК Вы можете назвать?



МикроРНК (англ. microRNA, miRNA) — малые некодирующие молекулы РНК длиной 18—25 нуклеотидов (в среднем 22), принимающие участие в транскрипционной и посттранскрипционной регуляции экспрессии генов путём РНК-интерференции.

**А какие еще классы некодирующих РНК Вы можете назвать?**

1. snoRNA - Small nucleolar RNAs
2. siRNA - Small (short) interfering RNA
3. piRNA - Piwi-interacting RNA
4. circRNA - Circular RNA
5. shRNA - short (small ) hairpin RNA
6. tRNA - transfer RNA
7. lncRNA - Long non-coding RNA
8. vlincRNA – very long non-coding RNA





# Базы данных, содержащие информацию по метаболитам.



ChEBI

PubChem



- ChEBI
- PubChem
- HMDB
- И другие

# ChEBI

- Доступна через интернет по адресу <https://www.ebi.ac.uk/chebi/init.do>
- Содержит информацию по малым химическим соединениям, как природного, так и искусственного происхождения.
- Идентификаторы: числа. Например, ацетон имеет идентификатор 15347.
- Содержит информацию по
  - ✓ функциям
  - ✓ Химической формуле
  - ✓ Молекулярному весу
  - ✓ Биологической роли и др.

## Информационная карточка метаболита.

The screenshot shows the ChEBI website interface. At the top, there is a search bar with the text "CHEBI:153" and a "Search" button. Below the search bar, there are navigation links: Home, Advanced Search, Browse, Documentation, Download, Tools, About ChEBI, Preferences, and Submit. The main content area displays the title "CHEBI:15347 - acetone" and a tabbed interface with "Main" selected. The "Main" tab shows the chemical structure of acetone (H<sub>3</sub>C-C(=O)-CH<sub>3</sub>) and a table of properties:

|                      |  |
|----------------------|--|
| ChEBI Name           | acetone  |
| ChEBI ID             | CHEBI:15347  |
| Definition           | A methyl ketone that consists of propane bearing an oxo group at C2. |
| Stars                | ☆☆☆ This entity has been manually annotated by the ChEBI Team.       |
| Secondary ChEBI IDs  | CHEBI:40571, CHEBI:2398, CHEBI:13708, CHEBI:22182                    |
| Supplier Information | ZINC00895111, eMolecules:474422                                      |

Below the table, there are links for "Download Molfile", "Find compounds which contain this structure", "Find compounds which resemble this structure", and "Take structure to the Advanced Search". At the bottom, there is a "Wikipedia" section with a "License" button and a paragraph of text describing acetone as a colorless, volatile, flammable liquid used as a solvent in laboratories.

# PubChem

- Доступна через интернет по адресу <https://pubchem.ncbi.nlm.nih.gov/compound/>
- Содержит информацию по малым химическим соединениям, как природного, так и искусственного происхождения.
- Идентификаторы: числа. Например, ацетон имеет идентификатор 180.
- Содержит информацию по
  - ✓ синонимам
  - ✓ функциям и физико-химическим свойствам
  - ✓ Химической формуле
  - ✓ Молекулярному весу
  - ✓ Биологической роли и др.

## Информационная карточка метаболита.

Надежный | <https://pubchem.ncbi.nlm.nih.gov/compound/180#section=Top>

NIH | NLM | National Center for Biotechnology Information

PubChem | OPEN CHEMISTRY DATABASE

Search Compounds

Compound Summary for CID 180

Download | Share | Help

### Acetone

Cite this Record

STRUCTURE | VENDORS | PHARMACOLOGY | LITERATURE | PATENTS | BIOACTIVITIES

|                     |  |
|---------------------|--|
| PubChem CID:        | 180  |
| Chemical Names:     | Acetone; 2-propanone; Propanone; Dimethyl ketone; Methyl ketone; 67-64-1 <a href="#">More...</a>             |
| Molecular Formula:  | C <sub>3</sub> H <sub>6</sub> O or CH <sub>3</sub> -CO-CH <sub>3</sub> or (CH <sub>3</sub> ) <sub>2</sub> CO |
| Molecular Weight:   | 58.08 g/mol  |
| InChI Key:          | CSCPPACGZOO CGX-UHFFFAOYSA-N   |
| Substance Registry: | FDA UNII   |
| Safety Summary:     | <a href="#">Laboratory Chemical Safety Summary (LCSS)</a>  |

Acetone is a colorless liquid used as a solvent and an antiseptic. It is one of the ketone bodies produced during ketoacidosis. [from MeSH](#)

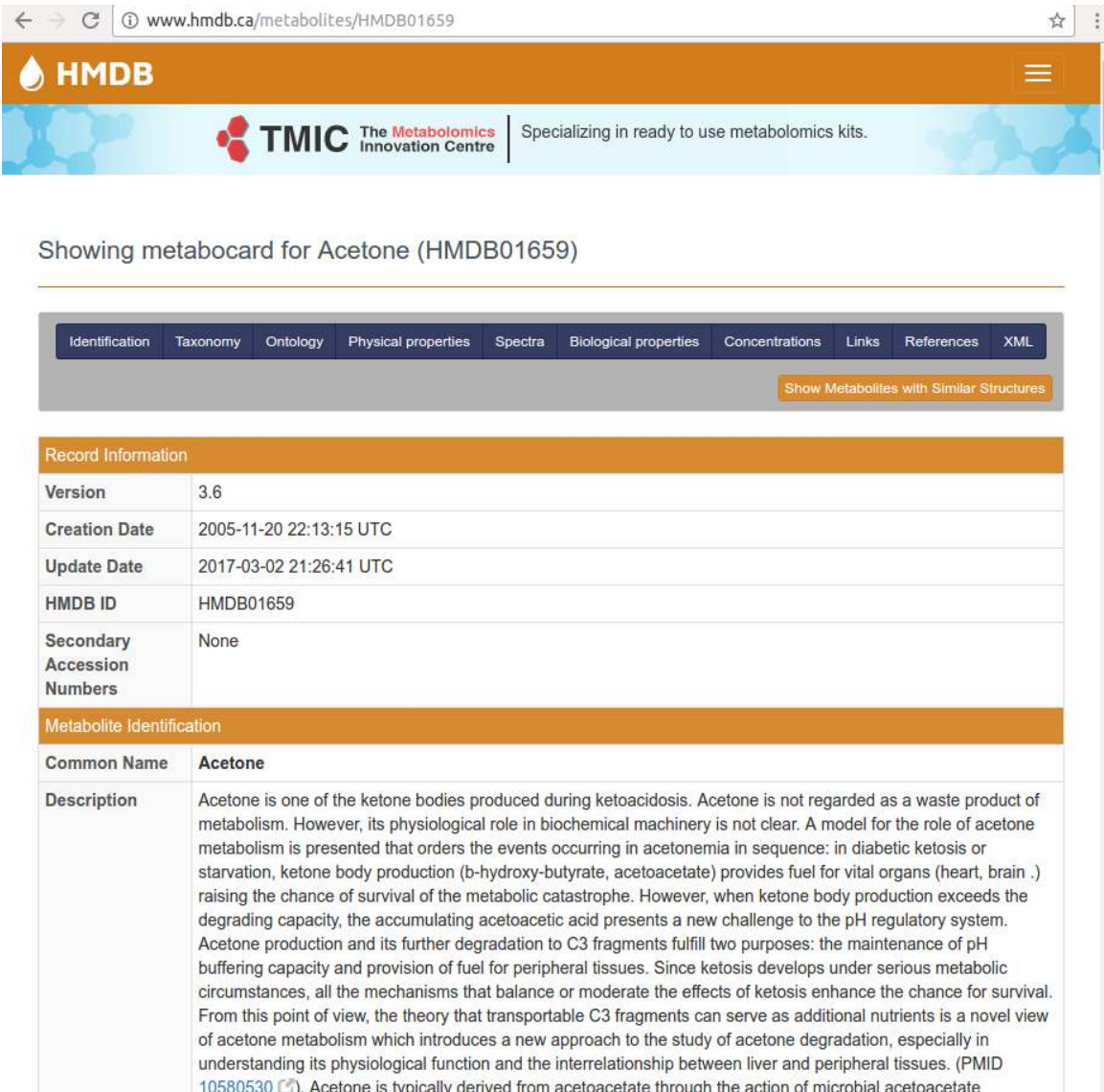
Acetone is a colorless, volatile, flammable organic solvent. Acetone occurs naturally in plants, trees, forest fires, vehicle exhaust and as a breakdown product of animal fat metabolism. This agent may be normally present in very small quantities in urine and blood; larger amounts may be found in the urine and blood of diabetics. Acetone is toxic in high doses. (NCI04) [Pharmacology from NCI](#)

Acetone is one of the ketone bodies produced during ketoacidosis. Acetone is not regarded as a waste product of metabolism. However, its physiological role in biochemical machinery is not clear. A model for the role of acetone metabolism is presented that orders the events occurring in acetonemia in sequence: in diabetic ketosis or starvation, ketone body production (β-hydroxy-butyrate, acetoacetate) provides fuel for vital organs (heart, brain . . .) raising the chance of survival of the metabolic catastrophe. However, when ketone body production exceeds the degrading capacity, the accumulating acetoacetic acid presents a new challenge to the pH regulatory system. Acetone production and its further degradation to C<sub>3</sub> fragments fulfill two purposes: the maintenance of pH buffering capacity and provision of fuel for peripheral tissues. Since ketosis develops under serious metabolic circumstances, all the mechanisms that balance or moderate the effects of ketosis enhance the chance for survival. From this point of view, the theory that transportable C<sub>3</sub> fragments can serve as

# HMDB

- Доступна через интернет по адресу <http://www.hmdb.ca/>
- Содержит информацию по метаболитам человека.
- Идентификаторы: HMDB затем пять чисел. Например, ацетон имеет идентификатор HMDB01659.
- Содержит информацию по
  - ✓ синонимам
  - ✓ функциям и физико-химическим свойствам
  - ✓ Химической формуле, молекулярному весу
  - ✓ Ассоциированным заболеваниям
  - ✓ Концентрации в норме и при патологии в различных тканях
  - ✓ Биологической роли и др.

## Информационная карточка метаболита.



Showing metabocard for Acetone (HMDB01659)

Identification Taxonomy Ontology Physical properties Spectra Biological properties Concentrations Links References XML

Show Metabolites with Similar Structures

| Record Information          |                         |
|-----------------------------|-------------------------|
| Version                     | 3.6                     |
| Creation Date               | 2005-11-20 22:13:15 UTC |
| Update Date                 | 2017-03-02 21:26:41 UTC |
| HMDB ID                     | HMDB01659               |
| Secondary Accession Numbers | None                    |

| Metabolite Identification |  |
|---------------------------|--|
| Common Name               | Acetone  |
| Description               | Acetone is one of the ketone bodies produced during ketoacidosis. Acetone is not regarded as a waste product of metabolism. However, its physiological role in biochemical machinery is not clear. A model for the role of acetone metabolism is presented that orders the events occurring in acetonemia in sequence: in diabetic ketosis or starvation, ketone body production (b-hydroxy-butyrate, acetoacetate) provides fuel for vital organs (heart, brain .) raising the chance of survival of the metabolic catastrophe. However, when ketone body production exceeds the degrading capacity, the accumulating acetoacetic acid presents a new challenge to the pH regulatory system. Acetone production and its further degradation to C3 fragments fulfill two purposes: the maintenance of pH buffering capacity and provision of fuel for peripheral tissues. Since ketosis develops under serious metabolic circumstances, all the mechanisms that balance or moderate the effects of ketosis enhance the chance for survival. From this point of view, the theory that transportable C3 fragments can serve as additional nutrients is a novel view of acetone metabolism which introduces a new approach to the study of acetone degradation, especially in understanding its physiological function and the interrelationship between liver and peripheral tissues. (PMID 10580530). Acetone is typically derived from acetoacetate through the action of microbial acetoacetate |

Мировое производство этого соединения составляет более 6,9 миллионов тонн в год (по данным на 2012 г) и устойчиво растёт. Широко используется в различном производстве и является популярным растворителем.

Является естественным метаболитом, производимым организмами млекопитающих. Некоторое количество вещества выводится с выдыхаемым воздухом и выделениями кожи, некоторое — с мочой. В крови человека в норме содержится 1-2 мг/100 мл этого вещества, в суточном количестве мочи — 0,01-0,03 г. При нарушениях обмена веществ, например, при сахарном диабете, в моче и крови повышается содержание этого соединения.

**Что это за соединение?**

Мировое производство этого соединения составляет более 6,9 миллионов тонн в год (по данным на 2012 г) и устойчиво растёт. Широко используется в различных производствах и является популярным растворителем.

Является естественным метаболитом, производимым организмами млекопитающих. Некоторое количество вещества выводится с выдыхаемым воздухом, выделениями кожи и мочой. В крови человека в норме содержится 1-2 мг/100 мл этого вещества, в суточном количестве мочи — 0,01-0,03 г. При нарушениях обмена веществ, например, при сахарном диабете, в моче и крови повышается содержание этого соединения.

Что это за соединение? **Ацетон**  $\text{H}_3\text{C}-\overset{\text{O}}{\parallel}-\text{CH}_3$



# Базы данных, содержащие информацию по заболеваниям.



- OMIM
- ICD - International Classification of Diseases
- Disease Ontology
- И другие



# OMIM

- Доступна через интернет по адресу <https://www.omim.org/>
- Содержит информацию по наследуемым заболеваниям человека.
- Идентификаторы: MIM:число. Например, DIABETES MELLITUS, NONINSULIN-DEPENDENT имеет идентификатор MIM: 125853.
- Содержит информацию по
  - ✓ синонимам
  - ✓ ассоциациям с генами
  - ✓ Наследуемости
  - ✓ Биохимических особенностях
  - ✓ Патогенезе
  - ✓ Животным моделям и др.

## Информационная карточка заболевания.

diabetes

# 125853 ICD+

**DIABETES MELLITUS, NONINSULIN-DEPENDENT; NIDDM**

*Alternative titles; symbols*

**DIABETES MELLITUS, TYPE II; T2D**  
**NONINSULIN-DEPENDENT DIABETES MELLITUS**  
**MATURITY-ONSET DIABETES**

Other entities represented in this entry:

**INSULIN RESISTANCE, SUSCEPTIBILITY TO, INCLUDED**  
**DIABETES MELLITUS, TYPE 2, PROTECTION AGAINST, INCLUDED**

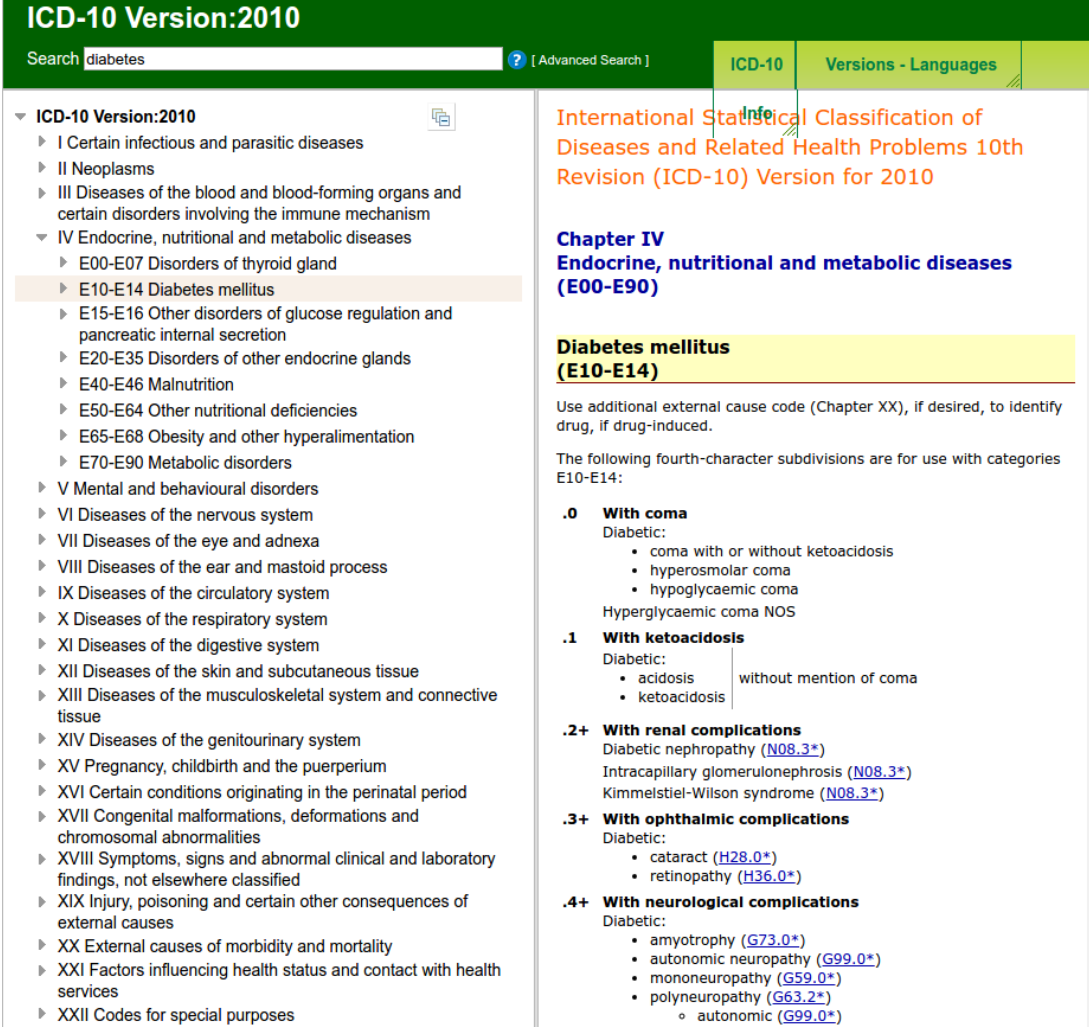
**Phenotype-Gene Relationships**

| Location | Phenotype                                 | Phenotype MIM number | Inheritance | Phenotype mapping key | Gene/Locus | Gene/Locus MIM number |
|----------|---|----------------------|-------------|-----------------------|------------|-----------------------|
| 2q24.1   | {Diabetes, type 2, susceptibility to}     | 125853               | AD          | 3                     | GPD2       | 138430                |
| 2q31.3   | {Diabetes mellitus, noninsulin-dependent} | 125853               | AD          | 3                     | NEUROD1    | 601724                |
| 2q36.3   | {Diabetes mellitus, noninsulin-dependent} | 125853               | AD          | 3                     | IRS1       | 147545                |
| 3p25.2   | {Diabetes, type 2}                        | 125853               | AD          | 3                     | PPARG      | 601487                |
| 3q26.2   | {Diabetes                                 | 125853               | AD          | 3                     | SLC2A2     | 138160                |



# ICD10 - International Statistical Classification of Diseases and Related Health Problems 10th Revision

- Доступна через интернет по адресу <http://apps.who.int/classifications/icd10/browse/2010>
- Содержит классификацию заболеваний человека, принятую международным сообществом и используемую практикующими врачами.
- Идентификаторы: заглавная буква одна из перечисленных - A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, V, Y, Z, U и далее двухзначное число, затем может следовать точка и число. Например, Diabetes mellitus это раздел с идентификаторами от E10 до E14. Заболевание Insulin-dependent diabetes mellitus имеет идентификатор E10, а postsurgical hypoinsulinaemia - E89.1.



The screenshot displays the ICD-10 Version:2010 web interface. The search bar contains 'diabetes'. The left sidebar shows a hierarchical tree of categories, with 'E10-E14 Diabetes mellitus' selected. The main content area shows the title 'International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) Version for 2010' and 'Chapter IV Endocrine, nutritional and metabolic diseases (E00-E90)'. Under 'Diabetes mellitus (E10-E14)', it provides instructions on using external cause codes and lists fourth-character subdivisions: .0 With coma, .1 With ketoacidosis, .2+ With renal complications, .3+ With ophthalmic complications, and .4+ With neurological complications.

**ICD-10 Version:2010**

Search  [ Advanced Search ]

ICD-10 Versions - Languages

**International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) Version for 2010**

**Chapter IV  
Endocrine, nutritional and metabolic diseases  
(E00-E90)**

**Diabetes mellitus  
(E10-E14)**

Use additional external cause code (Chapter XX), if desired, to identify drug, if drug-induced.

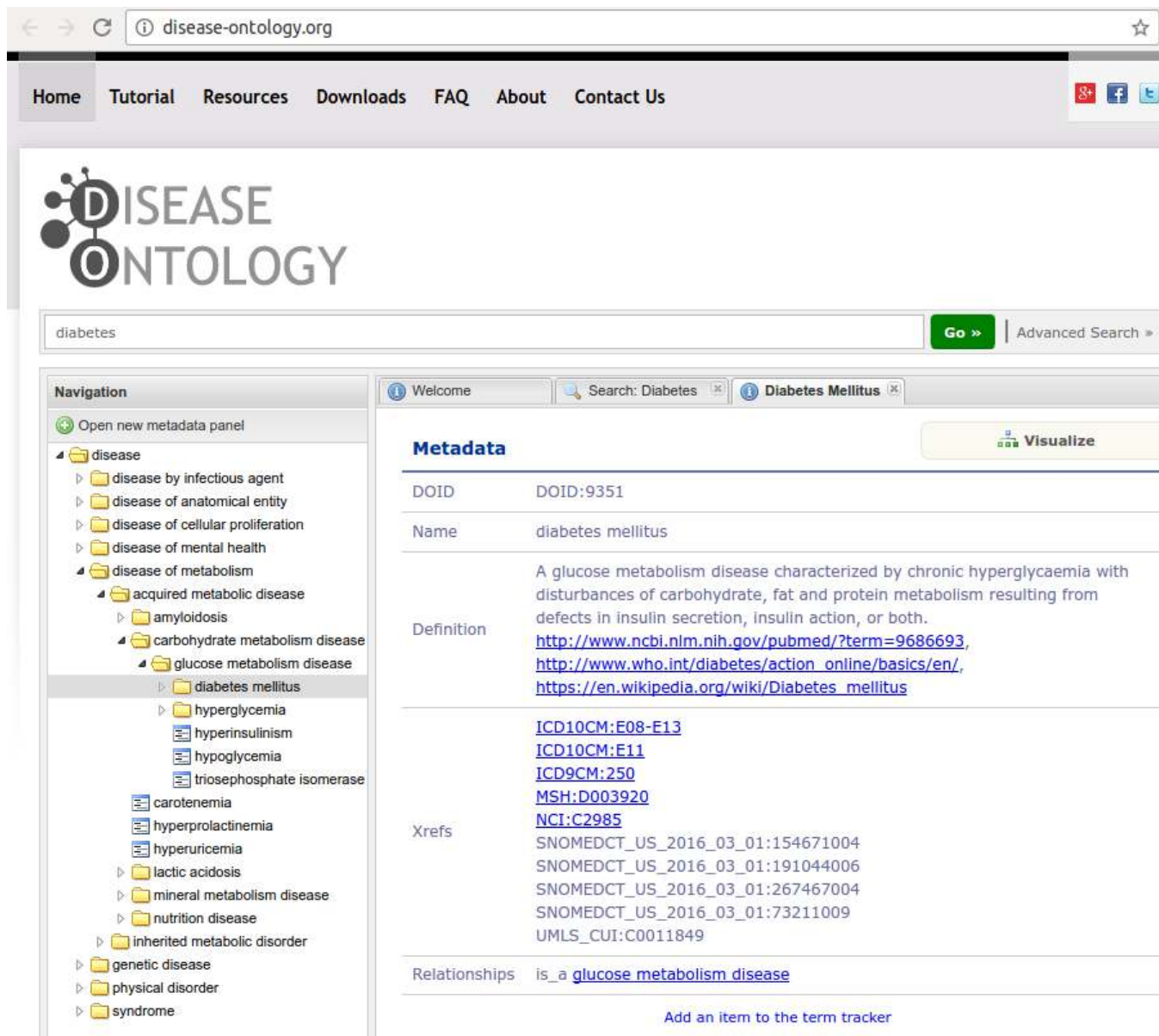
The following fourth-character subdivisions are for use with categories E10-E14:

- .0 With coma**  
Diabetic:
  - coma with or without ketoacidosis
  - hyperosmolar coma
  - hypoglycaemic comaHyperglycaemic coma NOS
- .1 With ketoacidosis**  
Diabetic:
  - acidosis without mention of coma
  - ketoacidosis
- .2+ With renal complications**  
Diabetic nephropathy (N08.3\*)  
Intracapillary glomerulonephrosis (N08.3\*)  
Kimmelstiel-Wilson syndrome (N08.3\*)
- .3+ With ophthalmic complications**  
Diabetic:
  - cataract (H28.0\*)
  - retinopathy (H36.0\*)
- .4+ With neurological complications**  
Diabetic:
  - amyotrophy (G73.0\*)
  - autonomic neuropathy (G99.0\*)
  - mononeuropathy (G59.0\*)
  - polyneuropathy (G63.2\*)
    - autonomic (G99.0\*)

# Disease Ontology

- Доступна через интернет по адресу <http://disease-ontology.org/>
- Содержит классификацию заболеваний человека по 8 основным группам.
- Идентификаторы DOID:число. Например, Diabetes mellitus это раздел с идентификаторами от E10 до E14. Заболевание diabetes mellitus имеет идентификатор DOID:9351 и относится к группе glucose metabolism disease -> carbohydrate metabolism disease -> acquired metabolic disease -> disease of metabolism.
- Содержит краткую информацию по патогенезу заболевания и ссылки на другие источники.

## Информационная карточка заболевания.



The screenshot shows the Disease Ontology website interface. At the top, there is a navigation menu with links for Home, Tutorial, Resources, Downloads, FAQ, About, and Contact Us. The main header features the "DISEASE ONTOLOGY" logo. A search bar contains the term "diabetes" with a "Go" button and an "Advanced Search" link. Below the search bar, there are tabs for "Welcome", "Search: Diabetes", and "Diabetes Mellitus". The "Diabetes Mellitus" tab is active, displaying a "Metadata" section with the following information:

| Metadata      |   |
|---------------|---|
| DOID          | DOID:9351   |
| Name          | diabetes mellitus   |
| Definition    | A glucose metabolism disease characterized by chronic hyperglycaemia with disturbances of carbohydrate, fat and protein metabolism resulting from defects in insulin secretion, insulin action, or both.<br><a href="http://www.ncbi.nlm.nih.gov/pubmed/?term=9686693">http://www.ncbi.nlm.nih.gov/pubmed/?term=9686693</a> ,<br><a href="http://www.who.int/diabetes/action_online/basics/en/">http://www.who.int/diabetes/action_online/basics/en/</a> ,<br><a href="https://en.wikipedia.org/wiki/Diabetes_mellitus">https://en.wikipedia.org/wiki/Diabetes_mellitus</a> |
| Xrefs         | <a href="#">ICD10CM:E08-E13</a><br><a href="#">ICD10CM:E11</a><br><a href="#">ICD9CM:250</a><br><a href="#">MSH:D003920</a><br><a href="#">NCI:C2985</a><br>SNOMEDCT_US_2016_03_01:154671004<br>SNOMEDCT_US_2016_03_01:191044006<br>SNOMEDCT_US_2016_03_01:267467004<br>SNOMEDCT_US_2016_03_01:73211009<br>UMLS_CUI:C0011849  |
| Relationships | is_a <a href="#">glucose metabolism disease</a>   |

At the bottom of the card, there is a link to "Add an item to the term tracker". On the left side of the page, a "Navigation" panel shows a tree structure of disease categories, with "diabetes mellitus" highlighted under "glucose metabolism disease".

Оцените, сколько в настоящее время известно заболеваний?



Оцените, сколько в настоящее время известно заболеваний?

В МКБ-10 описано порядка 10 тысяч заболеваний и синдромов.



# Базы данных, содержащие информацию по биологическим путям/процессам.



- Gene Ontology
- KEGG
- Reactome
- И другие

# Gene Ontology

- Доступна через интернет по адресу <http://www.geneontology.org/>
- Является онтологией, описывающей некоторые основные биологические единицы. Содержит информацию по биологическим процессам, клеточным компонентам и молекулярным функциям с привязкой к генам различных организмов.
- Идентификаторы: GO:семизначное число. Например, glucose metabolic process имеет идентификатор GO:0006006.
- Содержит информацию по
  - ✓ синонимам
  - ✓ описанию
  - ✓ ассоциациям с генами и др.

## Информационная карточка биологического процесса.

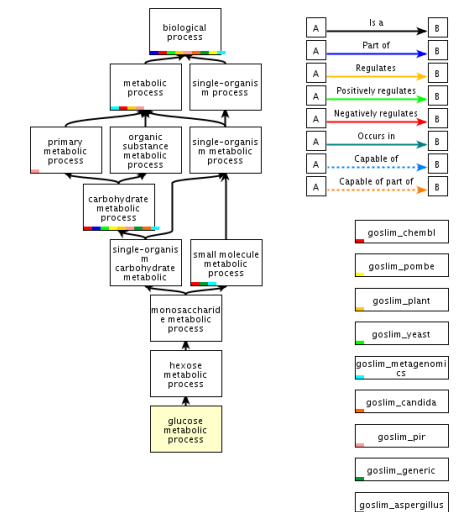
glucose metabolic process

### Term Information ⓘ

**Accession** GO:0006006 [Data health](#) ▼  
**Name** glucose metabolic process  
**Ontology** biological\_process  
**Synonyms** cellular glucose metabolic process, glucose metabolism  
**Alternate IDs** None  
**Definition** The chemical reactions and pathways involving glucose, the aldohexose gluco-hexose. D-glucose is dextrorotatory and is sometimes known as dextrose; it is an important source of energy for living organisms and is found free as well as combined in homo- and hetero-oligosaccharides and polysaccharides. *Source: ISBN:0198506732*  
**Comment** None  
**History** See term [history](#) for GO:0006006 at QuickGO  
**Subset** gosubset\_prok  
**Related**  
[Link](#) to all **genes and gene products** annotated to glucose metabolic process.  
[Link](#) to all direct and indirect **annotations** to glucose metabolic process.  
[Link](#) to all direct and indirect **annotations download** (limited to first 10,000) for glucose metabolic process.

| Gene/product             | Gene/product name | Annotation qualifier          | GO class (direct)             | Annotation extension | Contributor | Organism                    | Evidence |
|--------------------------|-------------------|-------------------------------|-------------------------------|----------------------|-------------|-----------------------------|----------|
| <input type="checkbox"/> | LOC101076683      | Uncharacterized protein       | glucose metabolic process     |                      | GO_Central  | Takifugu rubripes           | IBA      |
| <input type="checkbox"/> | Soly03g118650.2   |                               | glucose metabolic process     |                      | GO_Central  | Solanum lycopersicum        | IBA      |
| <input type="checkbox"/> | Os12g0190000      | Os12g0190000 protein          | glucose metabolic process     |                      | GO_Central  | Oryza sativa Japonica Group | IBA      |
| <input type="checkbox"/> | LOC101072156      | Uncharacterized protein       | gluconeogenesis               |                      | GO_Central  | Takifugu rubripes           | IBA      |
| <input type="checkbox"/> | SDHAF3            |                               | regulation of gluconeogenesis |                      | GO_Central  | Macaca mulatta              | IBA      |
| <input type="checkbox"/> | g6pc3.1           | Uncharacterized protein       | gluconeogenesis               |                      | GO_Central  | Xenopus tropicalis          | IBA      |
| <input type="checkbox"/> | CHLREDRAFT_113717 | Predicted protein             | glucose metabolic process     |                      | GO_Central  | Chlamydomonas reinhardtii   | IBA      |
| <input type="checkbox"/> | gdpgp1            | GDP-D-glucose phosphorylase 1 | glucose metabolic process     |                      | GO_Central  | Xenopus tropicalis          | IBA      |
| <input type="checkbox"/> | FBP1              |                               | gluconeogenesis               |                      | GO_Central  | Pan troglodytes             | IBA      |
| <input type="checkbox"/> | G6PC2             | Uncharacterized protein       | gluconeogenesis               |                      | GO_Central  | Macaca mulatta              | IBA      |

View this term in QuickGO.



# KEGG - Kyoto Encyclopedia of Genes and Genomes

- Доступна через интернет по адресу <http://www.genome.jp/kegg/>
- Содержит описание биологических процессов на молекулярно-генетическом уровне, составленное вручную экспертами.
- Идентификаторы: трехбуквенный идентификатор организма и пятизначное число. Например, Glycolysis/Gluconeogenesis имеет идентификатор hsa00010.
- Содержит информацию по
  - ✓ Описанию
  - ✓ Входящим в процесс генам/белкам, метаболитам и др.
  - ✓ Ассоциированным заболеваниями
  - ✓ ссылкам на литературу и др.

## Информационная карточка биологического процесса.

**KEGG** PATHWAY: hsa00010 Help

|                    |   |         |
|--------------------|---|---------|
| <b>Entry</b>       | hsa00010  | Pathway |
| <b>Name</b>        | Glycolysis / Gluconeogenesis - Homo sapiens (human)   |         |
| <b>Description</b> | Glycolysis is the process of converting glucose into pyruvate and generating small amounts of ATP (energy) and NADH (reducing power). It is a central pathway that produces important precursor metabolites: six-carbon compounds of glucose-6P and fructose-6P and three-carbon compounds of glyceraldehyde-3P, glyceraldehyde-3P, glyceraldehyde-3P, phosphoenolpyruvate, and pyruvate [MD:M00001]. Acetyl-CoA, another important precursor metabolite, is produced by oxidative decarboxylation of pyruvate [MD:M00307]. When the enzyme genes of this pathway are examined in completely sequenced genomes, the reaction steps of three-carbon compounds from glyceraldehyde-3P to pyruvate form a conserved core module [MD:M00002], which is found in almost all organisms and which sometimes contains operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway of glucose from noncarbohydrate precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths [MD:M00003]. |         |
| <b>Class</b>       | Metabolism; Carbohydrate metabolism<br><a href="#">BRITE hierarchy</a>  |         |
| <b>Pathway map</b> | hsa00010 Glycolysis / Gluconeogenesis   |         |

# Reactome

- Доступна через интернет по адресу <http://www.reactome.org/>
- Содержит описание биологических процессов на молекулярно-генетическом уровне, составленное вручную экспертами.
- Идентификаторы: код группы процессов - трехбуквенный идентификатор организма и пятизначное число. Например, Glucose metabolism имеет идентификатор R-HSA-70326.
- Содержит информацию по
  - ✓ Описанию
  - ✓ Входящим в процесс генам/белкам, метаболитам и др.
  - ✓ Положению в иерархической структуре процессов
  - ✓ «ортологичным» процессам и др.

## Информационная карточка биологического процесса.

The image shows a screenshot of the Reactome website. At the top, the Reactome logo is displayed with the tagline 'A CURATED PATHWAY DATABASE'. Below the logo is a navigation menu with links for 'About', 'Content', 'Documentation', 'Tools', 'Community', 'Download', and 'Contact'. A search bar contains the text 'e.g. O95631, N'. The main content area is titled 'Glucose metabolism' and includes a table with the following information:

|                   |              |
|-------------------|--------------|
| Stable Identifier | R-HSA-70326  |
| Type              | Pathway      |
| Species           | Homo sapiens |

Below the table, there is a section titled 'Locations in the PathwayBrowser'. The main part of the screenshot is a detailed pathway diagram showing various metabolic steps and molecules. The diagram is color-coded and includes labels for 'Digestion of dietary carbohydrate', 'Galactose catabolism', 'Glucose metabolism', 'Fructose metabolism', 'Ribose phosphate pathway (ribose monophosphate shunt)', and 'Catabolism of glucuronate to xylulose-5-phosphate'. On the right side of the diagram, there is a text box that reads: 'breakdown is a major source and conversion to glucose yield pyruvate, glycogen can be synthesized from p'. At the bottom of the diagram, there is a 'Description' section with the text: 'These pathways together are responsible for: 1) the extraction of energy and carbon skeletons for biosyntheses from dietary sugars and related molecules; 2) the short-term storage of glucose in the body (as glycogen) and its mobilization d synthesis of glucose from pyruvate during extended fasts.'



Какой процесс отвечает за «потерю хвостика» головастиком?



Какой процесс отвечает за «потерю хвостика» головастиком?

Апоптоз (др.-греч. ἀπόπτωσης — листопад) — регулируемый процесс программируемой клеточной гибели, в результате которого клетка распадается на отдельные апоптотические тельца, ограниченные плазматической мембраной.



### 3. Интеграция разрозненных источников информации.

- Интеграция предполагает:

- ✓ единый формат данных
- ✓ единый способ графического и текстового представления данных
- ✓ возможность выполнения поискового запроса ко всем данным

- Преимущества интеграции биологических данных:

- 1) Установление эквивалентности объектов из различных источников с учетом синонимии, что позволяет избежать дублирования информации
- 2) Выявление противоречий, ошибок, пробелов в информации
- 3) Возможность выявления наиболее достоверной информации
- 4) Максимальная формализация информации
- 5) Обеспечивается структурированный поиск информации во всем информационном пространстве
- 6) Возможность автоматического анализа информации
- 7) Снижается трудоемкость работы с большими объемами информации

**Интеграция биологических данных облегчает реконструкцию целостных моделей живых систем, а не только частных аспектов их функционирования.**

# Примеры систем, интегрирующих биологические данные.



**GeneCards** - интеграционный ресурс по генам ЧЕЛОВЕКА, включает информацию из более чем 125 интернет-доступных ресурсов.



**Ensembl** - геномный браузер для исследования геномов позвоночных 87 видов, интегрирует более ста ресурсов.



**UCSC Genome Browser** - геномный браузер для исследования геномов около 50 видов живых организмов, интегрирует более ста ресурсов.



**GEO** - репозиторий, в котором хранятся разнородные данные по экспрессии генов, полученные в 2 124 727 экспериментах.



**UniProt** - содержит информацию по структуре и функциям белков 457010 бактерий, 168308 вирусов, 12163 архей, 894013 эукариот, интегрирует информацию из 145 ресурсов.



**NCBI** – ресурс, который обеспечивает общее информационное пространство для 66 баз данных.

# Список баз данных, представленных на сайте NCBI (<https://www.ncbi.nlm.nih.gov/>).

|  |   |   |
|--|---|---|
| Assembly   | GeneReviews                                       | Probe                                   |
| BioProject (formerly Genome Project)                 | Genes and Disease                                 | Protein Clusters                        |
| BioSample  | Genetic Testing Registry (GTR)                    | <b>Protein Database</b>                 |
| BioSystems   | Genome  | <b>PubChem BioAssay</b>                 |
| Bookshelf  | Genome Reference Consortium (GRC)                 | <b>PubChem Compound</b>                 |
| ClinicalTrials.gov                                   | HIV-1   | <b>PubChem Substance</b>                |
| <b>ClinVar</b>                                       | HomoloGene  | <b>PubMed</b>                           |
| CloneDB (formerly Clone Registry)                    | Influenza Virus                                   | <b>PubMed Central (PMC)</b>             |
| Computational Resources from NCBI's Structure Group  | Journals in NCBI Databases                        | PubMed Health                           |
| Consensus CDS (CCDS)                                 | MedGen  | Reference Sequence (RefSeq)             |
| Conserved Domain Database (CDD)                      | MeSH Database                                     | <b>RefSeqGene</b>                       |
| Database of Expressed Sequence Tags (dbEST)          | National Library of Medicine (NLM) Catalog        | Retrovirus Resources                    |
| Database of Genome Survey Sequences (dbGSS)          | NCBI C++ Toolkit Manual                           | SARS CoV                                |
| Database of Genomic Structural Variation (dbVar)     | NCBI Education Page                               | Sequence Read Archive (SRA)             |
| Database of Genotypes and Phenotypes (dbGaP)         | NCBI Glossary                                     | Structure (Molecular Modeling Database) |
| Database of Major Histocompatibility Complex (dbMHC) | NCBI Handbook                                     | Taxonomy                                |
| <b>Database of Short Genetic Variations (dbSNP)</b>  | NCBI Help Manual                                  | Third Party Annotation (TPA) Database   |
| <b>GenBank</b>                                       | NCBI Pathogen Detection Project                   | Trace Archive                           |
| <b>Gene</b>  | NCBI Website Search                               | UniGene                                 |
| <b>Gene Expression Omnibus (GEO) Database</b>        | <b>Nucleotide Database</b>                        | UniGene Library Browser                 |
| <b>Gene Expression Omnibus (GEO) Datasets</b>        | <b>Online Mendelian Inheritance in Man (OMIM)</b> | Viral Genomes                           |
| <b>Gene Expression Omnibus (GEO) Profiles</b>        | PopSet  | Virus Variation                         |

# Результат поискового запроса к ресурсу NCBI по ключевому слову YY1

Search NCBI databases

[Help](#)

Results found in 33 databases for "yy1"

## Literature

|                                |       |   |
|--------------------------------|-------|---|
| <a href="#">Books</a>          | 54    | books and reports                               |
| <a href="#">MeSH</a>           | 14    | ontology used for PubMed indexing               |
| <a href="#">NLM Catalog</a>    | 0     | books, journals and more in the NLM Collections |
| <a href="#">PubMed</a>         | 1,274 | scientific & medical abstracts/citations        |
| <a href="#">PubMed Central</a> | 5,868 | full-text journal articles                      |

## Health

|                               |    |  |
|-------------------------------|----|--|
| <a href="#">ClinVar</a>       | 49 | human variations of clinical significance        |
| <a href="#">dbGaP</a>         | 3  | genotype/phenotype interaction studies           |
| <a href="#">GTR</a>           | 4  | genetic testing registry                         |
| <a href="#">MedGen</a>        | 0  | medical genetics literature and links            |
| <a href="#">OMIM</a>          | 49 | online mendelian inheritance in man              |
| <a href="#">PubMed Health</a> | 0  | clinical effectiveness, disease and drug reports |

## Genomes

|                            |       |   |
|----------------------------|-------|---|
| <a href="#">Assembly</a>   | 0     | genome assembly information                       |
| <a href="#">BioProject</a> | 64    | biological projects providing data to NCBI        |
| <a href="#">BioSample</a>  | 100   | descriptions of biological source materials       |
| <a href="#">Clone</a>      | 770   | genomic and cDNA clones                           |
| <a href="#">dbVar</a>      | 859   | genome structural variation studies               |
| <a href="#">Genome</a>     | 17    | genome sequencing projects by organism            |
| <a href="#">GSS</a>        | 5     | genome survey sequences                           |
| <a href="#">Nucleotide</a> | 3,104 | DNA and RNA sequences                             |
| <a href="#">Probe</a>      | 658   | sequence-based probes and primers                 |
| <a href="#">SNP</a>        | 6,011 | short genetic variations                          |
| <a href="#">SRA</a>        | 147   | high-throughput DNA and RNA sequence read archive |
| <a href="#">Taxonomy</a>   | 0     | taxonomic classification and nomenclature catalog |

## Genes

|                              |         |  |
|------------------------------|---------|--|
| <a href="#">EST</a>          | 730     | expressed sequence tag sequences                       |
| <a href="#">Gene</a>         | 1,465   | collected information about gene loci                  |
| <a href="#">GEO DataSets</a> | 235     | functional genomics studies                            |
| <a href="#">GEO Profiles</a> | 103,477 | gene expression and molecular abundance profiles       |
| <a href="#">HomoloGene</a>   | 10      | homologous gene sets for selected organisms            |
| <a href="#">PopSet</a>       | 32      | sequence sets from phylogenetic and population studies |
| <a href="#">UniGene</a>      | 104     | clusters of expressed transcripts                      |

## Proteins

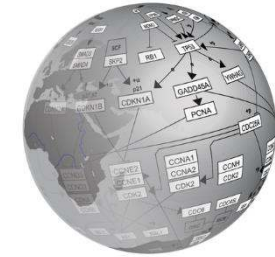
|                                   |     |   |
|-----------------------------------|-----|---|
| <a href="#">Conserved Domains</a> | 7   | conserved protein domains                         |
| <a href="#">Protein</a>           | 979 | protein sequences                                 |
| <a href="#">Protein Clusters</a>  | 1   | sequence similarity-based protein clusters        |
| <a href="#">Structure</a>         | 19  | experimentally-determined biomolecular structures |

## Chemicals

|                                   |     |  |
|-----------------------------------|-----|--|
| <a href="#">BioSystems</a>        | 767 | molecular pathways with links to genes, proteins and chemicals |
| <a href="#">PubChem BioAssay</a>  | 7   | bioactivity screening studies                                  |
| <a href="#">PubChem Compound</a>  | 1   | chemical information with structures, information and links    |
| <a href="#">PubChem Substance</a> | 280 | deposited substance and chemical information                   |

## 4. Ресурсы, интегрирующие биологическую информацию из разнородных источников и представляющие ее в виде генных сетей.

1. Ручная реконструкция генных сетей экспертами с использованием специальных программных средств – редакторов генных сетей.

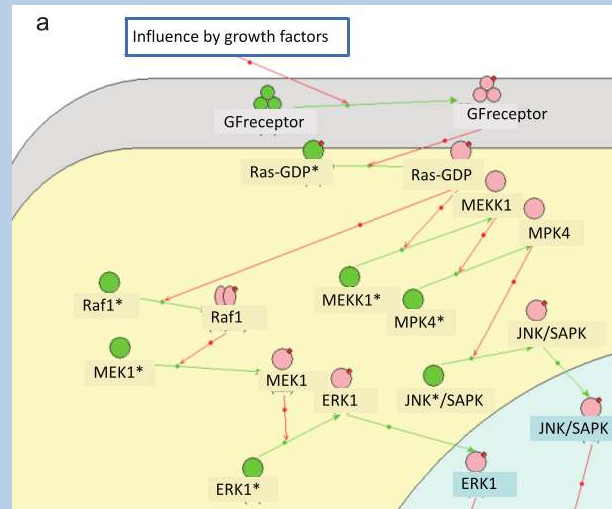


WIKIPATHWAYS  
Pathways for the People

SYSTEM FOR FORMALIZED DESCRIPTION,  
VISUALIZATION, AND MODELLING  
OF GENE NETWORKS



ИЦИГ СО РАН, 1998,  
N.A. Kolchanov, E.A.  
Ananko, N.L.  
Podkolodny, I.L.  
Stepanenko, E.V.  
Ignatieva, O.A.  
Podkolodnaya et al.



MAP-киназный путь  
передачи сигнала в  
ядро клетки,  
контролирующий  
процесс клеточного  
деления, активируе-  
мый ростовыми  
факторами.

2. Автоматическое извлечение знаний о молекулярно-генетических взаимодействиях различного типа из текстов научных публикаций и баз данных компьютерными методами (методы text-mining). Примерами являются ANDSystem, STRING, GeneMania, Pathway Commons и др.

# Система ANDSystem для автоматического извлечения знаний из текстов научных публикаций, международных патентов и баз данных в области биомедицины.

В ИЦиГ СО РАН разработана компьютерная технология автоматического извлечения знаний из текстов научных публикаций и международных патентов ANDSystem. Проведен автоматический анализ более 25 млн. научных публикаций и 10 млн. международных патентов.

Система ANDSystem включает

- модуль онтологии предметной области
- модуль текст майнинг
- базу знаний
- интерфейс пользователя.

Создана база знаний, содержащая более 15 млн. фактов, значимых для биомедицины, включающих взаимосвязи между

- молекулярно-генетическими системами и процессами
- заболеваниями и фенотипическими признаками
- факторами окружающей среды и др.

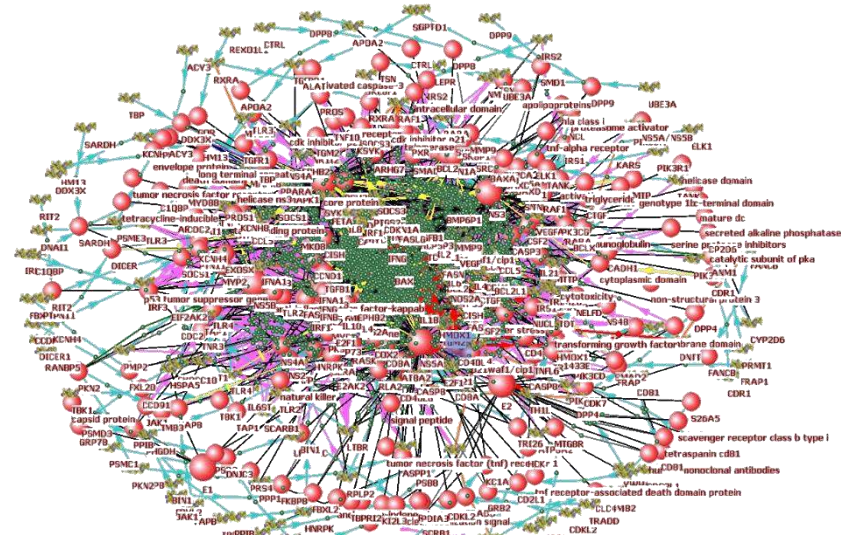


**Hepatitis C virus (HCV) NS5B protein is a membrane-associated phosphoprotein that possesses an RNA-dependent RNA polymerase activity. We recently reported that NS5A protein interacts with TRAF2 and modulates tumor necrosis factor alpha (TNF-alpha)-induced NF-kappaB and Jun N-terminal protein kinase (JNK). Since NS5A and NS5B are the essential components of the HCV replication complex, we examined whether NS5B could modulate TNF-alpha-induced NF-kappaB and JNK activation. In this study, we have demonstrated that TNF-alpha-induced NF-kappaB activation is inhibited by NS5B protein in HEK293 and hepatic cells. Furthermore, NS5B protein inhibited both TRAF2- and IKK-induced NF-kappaB activation. Using coimmunoprecipitation assays, we show that NS5B interacts with IKKalpha. Most importantly, NS5B protein in HCV subgenomic replicon cells interacted with endogenous IKKalpha, and men TNF-alpha-mediated IKKalpha kinase activation was significantly decreased by NS5B. Using in vitro kinase assay, we have further found that NS5B protein synergistically activated TNF-alpha-mediated JNK activity in HEK293 and hepatic cells. These data suggest that NS5B protein modulates TNF-alpha signaling pathway and may contribute to HCV pathogenesis.**

The network diagram shows interactions between several proteins: NS5A, NS5B, TRAF2, JNK, TNF-alpha, IKKalpha, and NK-kB. The legend indicates:
 

- Red circle: белки (proteins)
- Yellow arrow: усиление активности (activation)
- Green arrow: подавление активности (inhibition)
- Grey line: образование комплекса (complex formation)

 The diagram shows that NS5B inhibits TNF-alpha-induced JNK activation and TNF-alpha-induced IKKalpha activation. It also shows that NS5B interacts with IKKalpha.





# Pathway Commons



База данных содержит 69 498 биологических путей и доступна по адресу <http://www.pathwaycommons.org/>

Pathway Commons ресурс, интегрирующий информацию из различных баз данных. Биологические пути, представленные в Pathway Commons загружены непосредственно из исходных баз данных. Реконструкция биологических путей в исходных базах данных могла проводиться как путем ручного извлечения информации из литературы, так и путем компьютерной автоматической реконструкции. Качество биологических путей в Pathway Commons зависит от качества путей исходных баз данных. Pathway Commons позволяет пользователям фильтровать данные по различным критериям, включая источник информации.

## Pathway Commons интегрирует информацию из 24 баз данных

Reactome: 2007 pathways, 14427 interactions, 35835 participants  
NCI Pathway Interaction Database: Pathway: 745 pathways, 14707 interactions, 10531 participants  
PhosphoSitePlus: 27692 interactions, 15458 participants  
HumanCyc: 302 pathways, 7102 interactions, 5896 participants  
HPRD: 40595 interactions, 9844 participants  
PANTHER Pathway: 272 pathways, 4700 interactions, 6703 participants  
Database of Interacting Proteins: 8218 interactions, 4671 participants  
BioGRID: 322538 interactions, 645241 participants  
IntAct: 150549 interactions, 403729 participants  
BIND: 35279 interactions, 74675 participants  
CORUM: 4401 participants  
TRANSFAC: 427 pathways, 261624 interactions, 13276 participants  
miRTarBase: 5 pathways, 51214 interactions, 12775 participants  
DrugBank: 19297 interactions, 15854 participants  
Recon X: 1 pathways, 10813 interactions, 8316 participants  
Comparative Toxicogenomics Database: 32722 pathways, 390428 interactions, 61031 participants  
KEGG Pathway: 122 pathways, 3566 interactions, 3355 participants  
Small Molecule Pathway Database: 1206 pathways, 4701 interactions, 4863 participants  
Integrating Network Objects with Hierarchies: 774 pathways, 5432 interactions, 17142 participants  
NetPath: 27 pathways, 6347 interactions, 3266 participants  
WikiPathways: 333 pathways, 9758 interactions, 9584 participants  
ChEBI: All names  
SwissProt: All names  
UniChem: All names

Текстовое поле для ввода названия интересующего гена

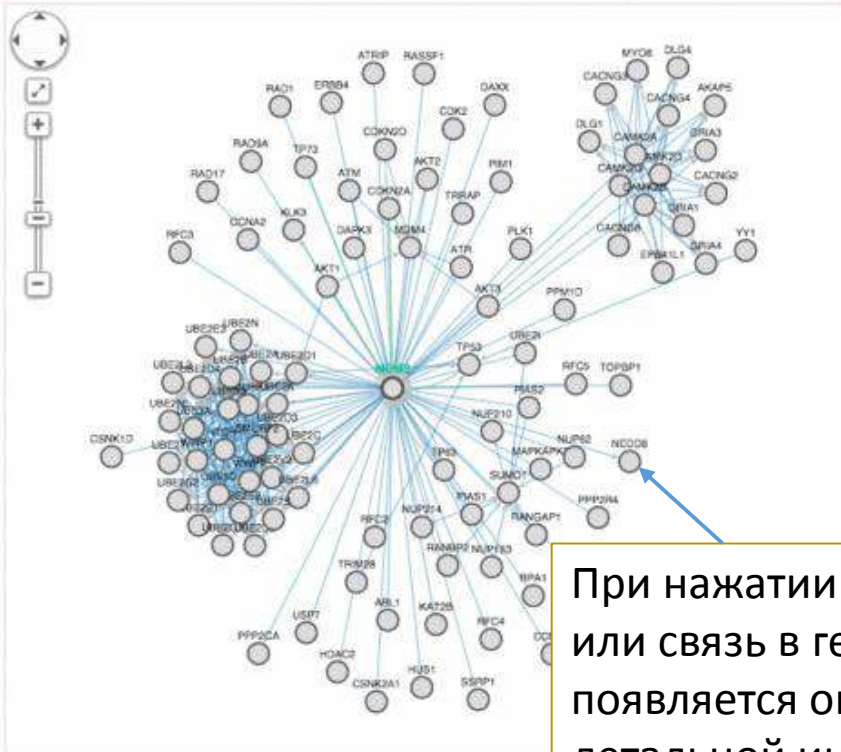
Добавление в сеть данных по изменениям, наблюдающимся при раке

Настройки для фильтрации сети по связям и вершинам

# PCViz Pathway Commons Network Visualizer

## Genes of interest

MDM2 +



Click on one of the interactions or genes in the network to details...

По умолчанию для заданного гена строится сеть, включающая всех соседей этого гена, однако, если у вас задано несколько генов, то можно просто увидеть взаимосвязи между ними.

Details Settings Context

Interaction types

- 362 controls state change X
- 409 controls expression X

Number of genes (468)

Slide-left to decrease the number of genes

Query type

Neighborhood

При нажатии на вершину или связь в генной сети появляется окно с более детальной информацией по данному объекту.

Details Settings Context

**CDT1**

The protein encoded by this gene is involved in the formation of the pre-replication complex that is necessary for DNA replication. The encoded protein can bind geminin, which prevents replication.

Aliases: DUP, R152

Description: chromatin licensing and DNA replication factor 1

Chromosome Location: 16q24.3

UniProt ID: Q9H211

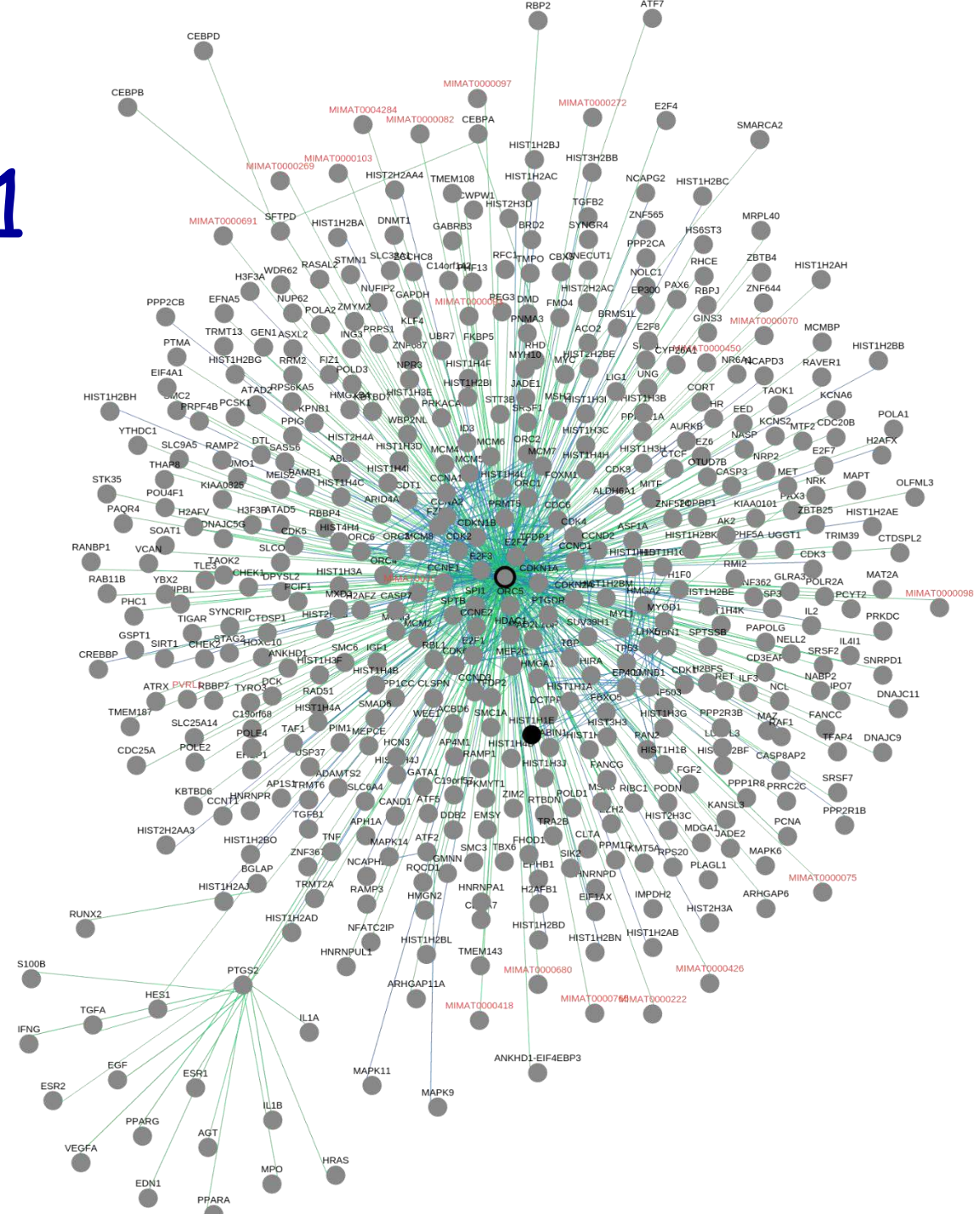
Gene ID: 81620

Добавление гена в список интересующих вас генов

# Пример генной сети для гена ретинобластомы RB1

Пример загрузки генной сети в формате SIF

|        |                 |        |
|--------|-----------------|--------|
| AATF   | in-complex-with | E2F1   |
| AATF   | in-complex-with | E2F2   |
| AATF   | in-complex-with | E2F3   |
| AATF   | in-complex-with | RB1    |
| AATF   | interacts-with  | RB1    |
| AATF   | in-complex-with | TFDP1  |
| ABCD3  | interacts-with  | DYRK1A |
| ABCD3  | interacts-with  | DYRK1B |
| ABCD3  | interacts-with  | RB1    |
| ABHD10 | interacts-with  | FOXK1  |



# 5. Примеры исследований, выполненных с использованием ресурсов ANDSystem, STRING, GeneMania, Pathway Commons.

ANDSystem

STRING

GeneMania

Pathway Commons

Системная биология и моделирование  
ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Валовский журнал генетики и селекции. 2016.20(6):840-847  
DOI 10.18699/VJ16.200

## SCIENTIFIC REPORTS

OPEN

Молекулярно-генетические механизмы взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при первичной открытоугольной глаукоме

O.V. Saik<sup>1</sup>, N.A. Kozovalova<sup>2</sup>, P.S. Demchenko<sup>3</sup>, N.V. Ivanitskiy<sup>3,4</sup>, T.V. Ivanitskiy<sup>3</sup>, A.E. Ivanitskiy<sup>3</sup>, M.N. Ponomareva<sup>5</sup>, O.S. Kozovalova<sup>2</sup>, O.A. Podkolodnaya<sup>2</sup>, I.N. Lazarev<sup>1,4</sup>, N.A. Kalchakov<sup>1</sup>, V.A. Ivanitskiy<sup>3,4</sup>

<sup>1</sup> Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Института цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия  
<sup>2</sup> Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный медицинский университет» Министерства здравоохранения Российской Федерации, Томск, Россия  
<sup>3</sup> Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия  
<sup>4</sup> Магистерский институт имени Отто фон Гуэрке, Магдебург, Германия

Received: 24 February 2016

Accepted: 10 June 2016

Published: 01 July 2016

## Alpha-tubulin enhanced renal tubular cell proliferation and tissue repair but reduced cell death and cell-crystal adhesion

Juthatip Manisom, Supaporn Khamchun, Arada Vainaphat & Visith Thongboonkerd

Adhesion of calcium oxalate (CaOx) crystals on renal tubular epithelial cells is a critical event for kidney stone disease that triggers many cascades of cellular response. Our previous expression proteomics study identified several altered proteins in MDCK renal tubular cells induced by CaOx crystals. However, functional significance of those changes had not been investigated. The present study thus aimed to define functional roles of such proteome data. Global protein network analysis using STRING software revealed  $\alpha$ -tubulin, which was decreased, as one of central nodes and its efficacy was confirmed. Overexpression of  $\alpha$ -tubulin (pcDNA6.2-TUBA1A) could maintain levels of  $\alpha$ -tubulin and its direct interacting partner, vimentin, after crystal exposure. Also, pcDNA6.2-TUBA1A successfully reduced cell death to almost the basal level and increased cell proliferation after crystal exposure. Additionally, tissue repair capacity was improved in pcDNA6.2-TUBA1A cells. Moreover, cell-crystal adhesion was reduced by pcDNA6.2-TUBA1A. Finally, levels of potential crystal receptors (HSP90, HSP70, and  $\alpha$ -anase) on apical membrane were dramatically reduced to basal levels by pcDNA6.2-TUBA1A. These findings implicate that  $\alpha$ -tubulin has protective roles in kidney stone disease by preventing cell death and cell-crystal adhesion, but on the other hand, enhancing cell proliferation and tissue repair function.

Until now, kidney stone disease is still a public health problem in almost all areas around the world. The disease causes substantial suffering and ultimately end-stage renal disease (ESRD). Unfortunately, the disease mechanisms remain poorly understood. Calcium oxalate (CaOx) is the major chemical component found in clinical stones<sup>1</sup>. This type of the stones can be originated from supersaturation of calcium and oxalate ions, leading to crystallization inside renal tubular fluid or urine<sup>2</sup>. CaOx crystals can then nucleate to form "stone nidus" and adhere directly onto apical surface of renal epithelial cells<sup>3,4</sup>. Adhesion of crystals onto the cells is a critical event, which triggers many cascades of cellular responses, e.g. cytotoxicity, injury, proliferation and apoptosis, that ultimately lead to kidney stone formation<sup>5,6</sup>. CaOx crystals also evoke inflammatory processes that can lead to fibrosis, loss of nephron and eventually ESRD<sup>7</sup>.

Even with the aforementioned knowledge, molecular mechanisms of the downstream cellular response remain largely unknown. From our previous expression proteomics study<sup>8</sup>, we have identified a number of proteins with altered levels in MDCK renal tubular cells in response to CaOx crystals. Those altered proteins were involved in various biological processes, i.e. ubiquitination pathway, signal transduction, cellular structure, purine biosynthesis, metabolic enzyme, retinol biosynthesis, cellular transportation, protein degradation, RNA metabolism, RNA binding protein, cell surface antigen, nucleic acid metabolism, antioxidant enzyme, chaperone, carrier protein, and protein biosynthesis. However, functional significance of those altered proteins had not been investigated. In the present study, we thus performed global protein network analysis of those altered proteins. Subsequently, overexpression of a protein, which was one of the central nodes of such protein-protein interactions network, was performed. Moreover, functional investigations were performed to address functional significance of the central node protein and its associated partners in kidney stone disease.

Medical Proteomics Unit, Office for Research and Development, Faculty of Medicine Siriraj Hospital, and Center for Research in Complex Systems Science, Mahidol University, Bangkok 10700, Thailand. Correspondence and requests for materials should be addressed to V.T. (email: thongboonkerd@id.com)

RESEARCH

Open Access



## Gene expression network analyses in response to air pollution exposures in the trucking industry

Jen-hwa Chu<sup>1</sup>\*, Jaime E. Hart<sup>2,3</sup>, Divya Chhabra<sup>2</sup>, Eric Garshick<sup>2,4</sup>, Benjamin A. Raby<sup>2,5</sup> and Francine Laden<sup>2,3,6</sup>

### Abstract

**Background:** Exposure to air pollution, including traffic-related pollutants, has been associated with a variety of adverse health outcomes, including increased cardiopulmonary morbidity and mortality, and increased lung cancer risk.

**Methods:** To better understand the cellular responses induced by air pollution exposures, we performed genome-wide gene expression microarray analysis using whole blood RNA sampled at three time-points across the work weeks of 63 non-smoking employees at 10 trucking terminals in the northeastern US. We defined genes and gene networks that were differentially activated in response to  $PM_{2.5}$  (particulate matter  $\leq 2.5$  microns in diameter) and elemental carbon (EC) and organic carbon (OC).

**Results:** Multiple transcripts were strongly associated ( $p_{adj} < 0.001$ ) with pollutant levels (48, 260, and 49 transcripts for EC, OC, and  $PM_{2.5}$ , respectively), including 63 that were statistically significantly correlated with at least two out of the three exposures. These genes included many that have been implicated in ischemic heart disease, chronic obstructive pulmonary disease (COPD), lung cancer, and other pollution-related illnesses. Through the combination of Gene Set Enrichment Analysis and network analysis (using GeneMANIA), we identified a core set of 25 interrelated genes that were common to all three exposure measures and were differentially expressed in two previous studies assessing gene expression attributable to air pollution. Many of these are members of fundamental cancer-related pathways, including those related to DNA and metal binding, and regulation of apoptosis and also but include genes implicated in chronic heart and lung diseases.

**Conclusions:** These data provide a molecular link between the associations of air pollution exposures with health effects.

**Keywords:** Air pollution, Trucking industry, Gene expression, Network analysis

### Background

Air pollution exposures, have been associated with a number of adverse health effects, including greater morbidity and mortality risks for cardiopulmonary diseases, and increased risk of lung cancer [1–6]. However, the underlying biological mechanisms have not been fully elucidated. Human studies of global changes in gene expression following controlled exposures [7], or using *in vivo* models [8, 9] have provided some insights in this

regard, yet few studies have rigorously assessed the impact of air pollution on gene expression in real-life settings. For example, though observational studies have been conducted in individuals from geographic regions with differing levels of air pollution have suggested associations, [10] studies with more refined exposure measures have not been performed.

In this study, we characterized the cellular response induced by traffic-related air pollution exposures in a population of non-smoking US trucking industry employees. We performed genome-wide gene expression microarray analysis using whole blood RNA sampled at three time-points during the work week. We integrate

\* Correspondence: jen-hwa.chu@yale.edu

<sup>1</sup>Section of Pulmonary, Critical Care and Sleep Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA  
Full list of author information is available at the end of the article



© The Author(s). 2016 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

ISSN 2026-1191/16/28008-8  
Published online: 01 July 2016  
© ANDSystem 2016

© saik-1@mail.bionet.ru

SCIENTIFIC REPORTS | 6:28008 | DOI: 10.1038/srep28008

1

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

## A Multi-Method Approach for Proteomic Network Inference in 11 Human Cancers

Yasin Şenbabaoğlu<sup>1\*</sup>, Selçuk Onur Sümer<sup>1</sup>, Francisco Sánchez-Vega<sup>1</sup>, Debra Bemis<sup>1</sup>, Giovanni Ciriello<sup>1\*</sup>, Nikolaus Schultz<sup>2</sup>, Chris Sander<sup>1\*</sup>

<sup>1</sup> Computational Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America, <sup>2</sup> Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America

\* Current address: Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland \* senbabay@mskcc.org (YS); chris@sanderlab.org (CS)

### Abstract

Protein expression and post-translational modification levels are tightly regulated in neoplastic cells to maintain cellular processes known as "cancer hallmarks". The first Pan-Cancer initiative of The Cancer Genome Atlas (TCGA) Research Network has aggregated protein expression profiles for 3,467 patient samples from 11 tumor types using the antibody based reverse phase protein array (RPPA) technology. The resultant proteomic data can be utilized to computationally infer protein-protein interaction (PPI) networks and to study the commonalities and differences across tumor types. In this study, we compare the performance of 13 established network inference methods in their capacity to retrieve the curated Pathway Commons interactions from RPPA data. We observe that no single method has the best performance in all tumor types, but a group of six methods, including diverse techniques such as correlation, mutual information, and regression, consistently rank highly among the tested methods. We utilize the high performing methods to obtain a consensus network; and identify four robust and densely connected modules that reveal biological processes as well as suggest antibody-related technical biases. Mapping the consensus network interactions to Reactome gene lists confirms the pan-cancer importance of signal transduction pathways, innate and adaptive immune signaling, cell cycle, metabolism, and DNA repair; and also suggests several biological processes that may be specific to a subset of tumor types. Our results illustrate the utility of the RPPA platform as a tool to study proteomic networks in cancer.

### Author Summary

Pan-cancer proteomic datasets from The Cancer Genome Atlas provide a unique opportunity to study the functions of proteins in human cancers. Such datasets, where proteins are measured in different conditions and where correlations are informative, can enable the discovery of potentially causal protein-protein interactions, which may in turn shed light on the function of proteins. However, it has been shown that the dominant correlations in



OPEN ACCESS

**Citation:** Şenbabaoğlu Y, Sümer SO, Sánchez-Vega F, Bemis D, Ciriello G, Schultz N, et al. (2016) A Multi-Method Approach for Proteomic Network Inference in 11 Human Cancers. *PLoS Comput Biol* 12(7): e1004765. doi:10.1371/journal.pcbi.1004765

**Editor:** Christian von Meiring, University of Zurich and Swiss Institute of Bioinformatics, SWITZERLAND

**Received:** June 7, 2015

**Accepted:** January 20, 2016

**Published:** February 29, 2016

**Copyright:** © 2016 Şenbabaoğlu et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The pan-cancer RPPA dataset is available from The Cancer Proteome Atlas at [http://app1.usfcm.fhcrc.org/index.php?option=com\\_content&view=article&id=104&Itemid=104](http://app1.usfcm.fhcrc.org/index.php?option=com_content&view=article&id=104&Itemid=104).

**Funding:** CS and YS were supported by the National Institute of General Medical Science (P41GM103504) and the National Cancer Institute (U2CA143840). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

PLOS Computational Biology | DOI:10.1371/journal.pcbi.1004765 February 29, 2016

1/31

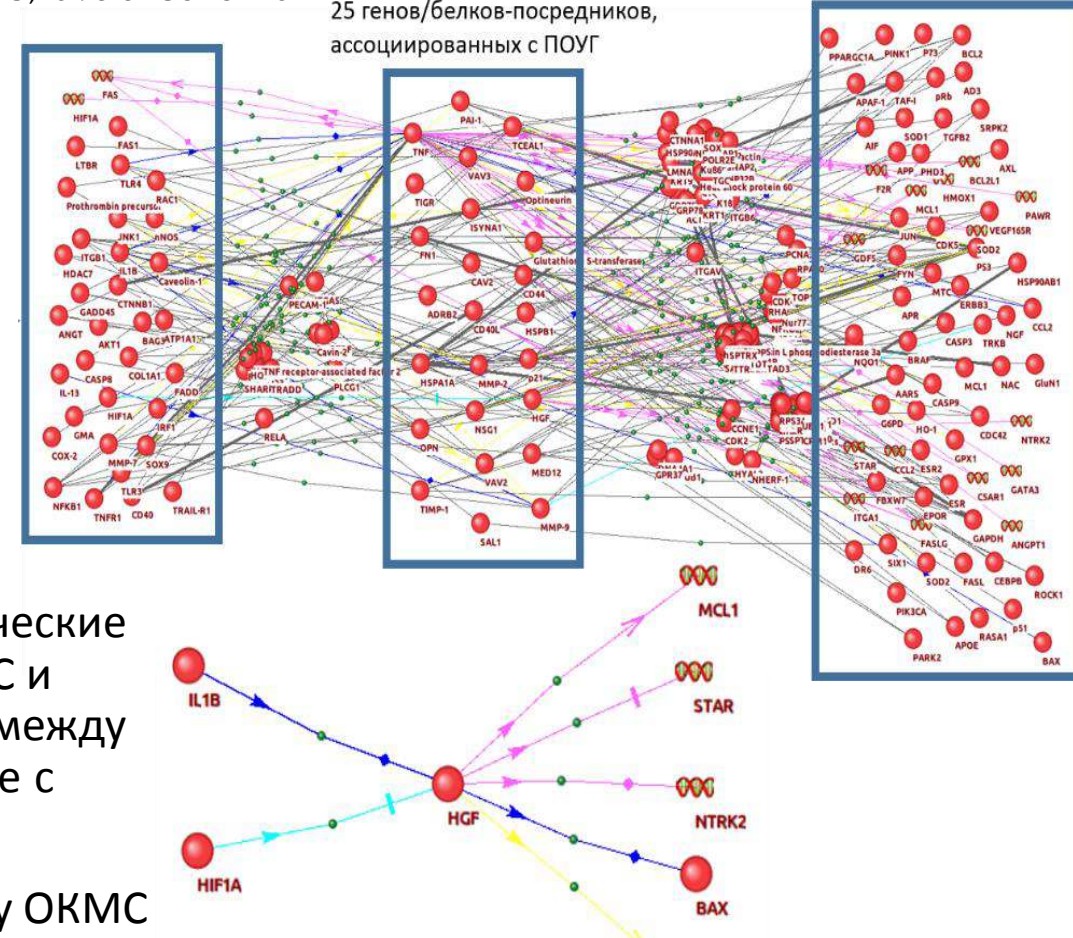
# Реконструкция молекулярно-генетических механизмов взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при первичной открыто-угольной глаукоме с помощью системы ANDSystem.

- Одной из основных причин необратимой потери зрения является первичная открыто-угольная глаукома (ПОУГ). Гибель клеток зрительного нерва может провоцироваться механическим стрессом, вызванным повышенным внутриглазным давлением, наблюдающимся при ПОУГ, индуцирующим нейрональный апоптоз.
- Целью работы был анализ ассоциативных сетей, описывающих молекулярно-генетические взаимодействия между белками и генами, вовлеченными в ответ клетки на механический стресс (ОКМС), нейрональный апоптоз и патогенез ПОУГ.
- Анализ ассоциативных сетей, описывающих молекулярно-генетические взаимодействия между белками и генами, вовлеченными в ОКМС и нейрональный апоптоз показал, что эти процессы тесно связаны между собой, а их участники имеют статистически значимое пересечение с генами, ассоциированными с ПОУГ.
- Можно предположить, что нарушения во взаимодействиях между ОКМС и нейрональным апоптозом, вызванные нарушением функцией генов, ассоциированных с ПОУГ, могут играть важную роль в патогенезе заболевания. Результаты анализа могут быть использованы для поиска новых лекарственных мишеней для терапии ПОУГ.

Белки/гены, вовлеченные в ОКМС

25 генов/белков-посредников, ассоциированных с ПОУГ

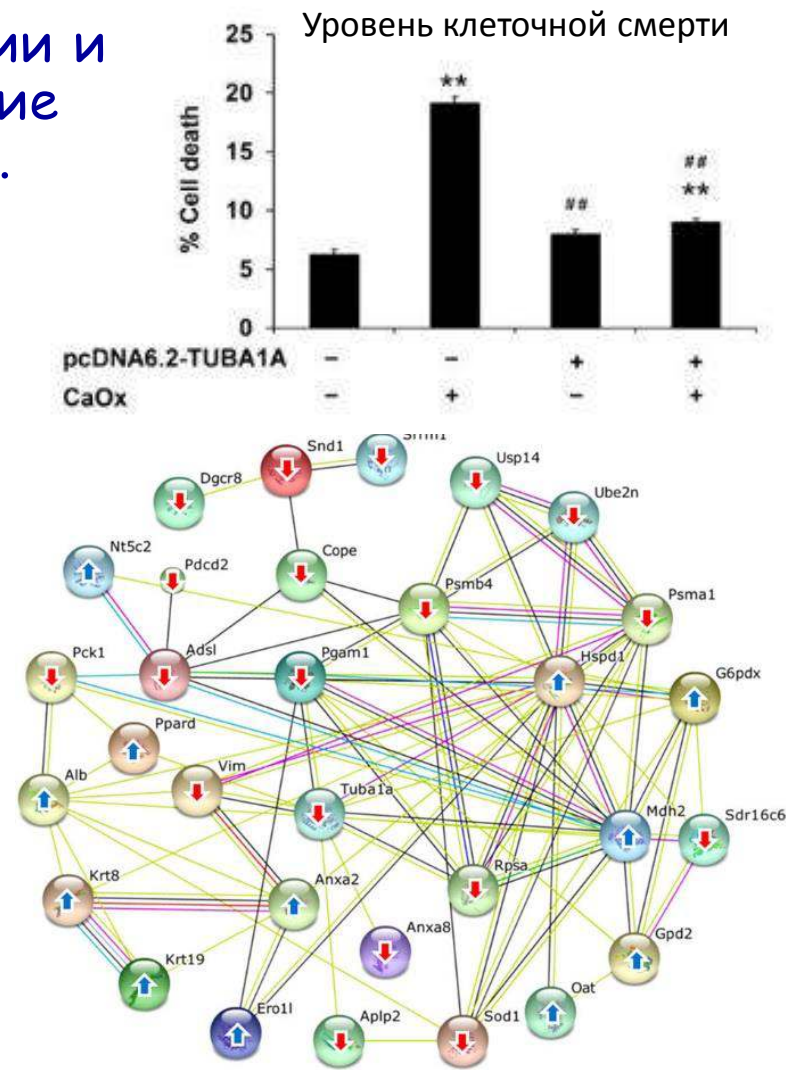
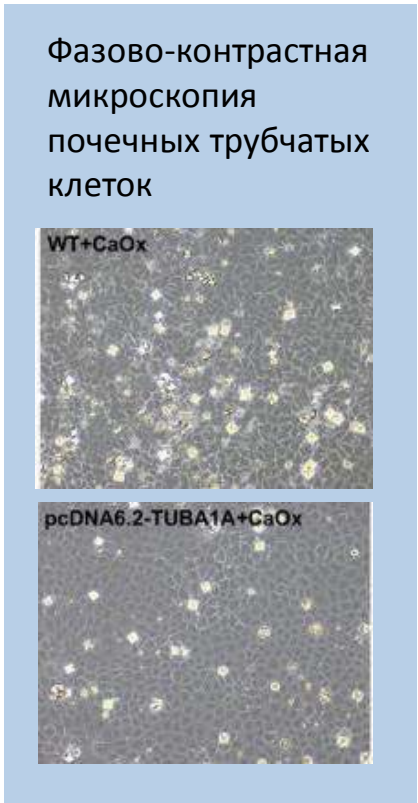
Белки/гены, вовлеченные в нейрональный апоптоз



Ассоциативная сеть, описывающая потенциальные взаимодействия между ОКМС и нейрональным апоптозом через молекулярно-генетические цепочки. а) Общая сеть, включающая 25 генов/белков-посредников, ассоциированных с ПОУГ. б) Пример цепочек, проходящих через фактор роста гепатоцитов (HGF), выступающий в качестве посредника, ассоциированного с ПОУГ.

## Способность альфа-тубулина усиливать процессы пролиферации и восстановления тканей и снижать клеточную смерть и связывание кристаллов оксалата кальция почечными трубчатыми клетками.

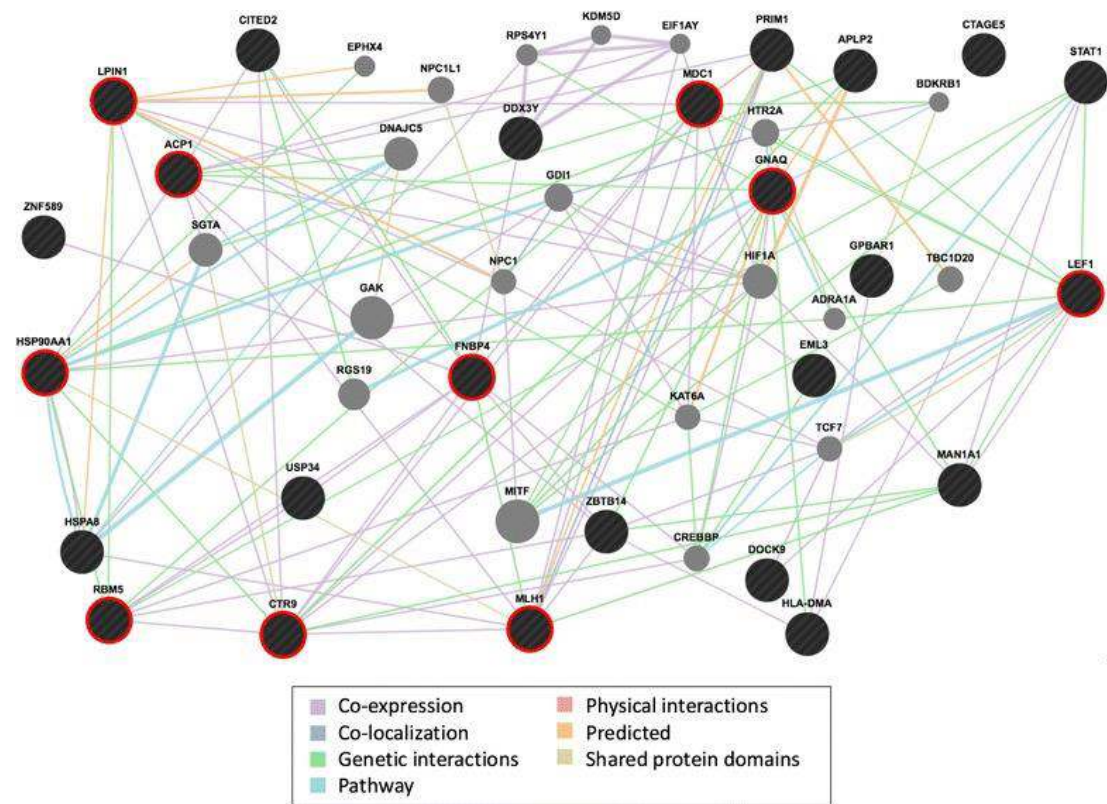
Адгезия кристаллов оксалата кальция (CaOx) на почечных канальцевых эпителиальных клетках наблюдается при мочекаменной болезни. Было показано, что экспрессия ряда белков достоверно меняется в клетках, индуцированных кристаллами CaOx. Анализ глобальной сети белок-белок взаимодействий с использованием системы STRING показал, что  $\alpha$ -тубулин, экспрессия которого была снижена в эксперименте, является одним из центральных узлов сети белок-белковых взаимодействий. Для подтверждения ключевой роли  $\alpha$ -тубулина была исследована клеточная линия pcDNA6.2-TUBA1A, в которой  $\alpha$ -тубулин сверхэкспрессирован. Было показано, что для клеток pcDNA6.2-TUBA1A наблюдается снижение гибели и увеличение клеточной пролиферации после воздействия кристаллов CaOx. Кроме того, в клетках pcDNA6.2-TUBA1A была улучшена репаративная способность тканей и снижена клеточно-кристаллическая адгезия. Эти результаты говорят в пользу того, что  $\alpha$ -тубулин играет ключевую роль в патогенезе почечнокаменной болезни.



Сеть взаимодействий между белками с измененной экспрессией в почечных трубчатых клетках MDCK, индуцированных кристаллами CaOx, реконструированная с помощью системы STRING.

## Анализ сети генов, экспрессия которых изменяется в ответ на воздействие загрязнения воздуха в автотранспортной отрасли.

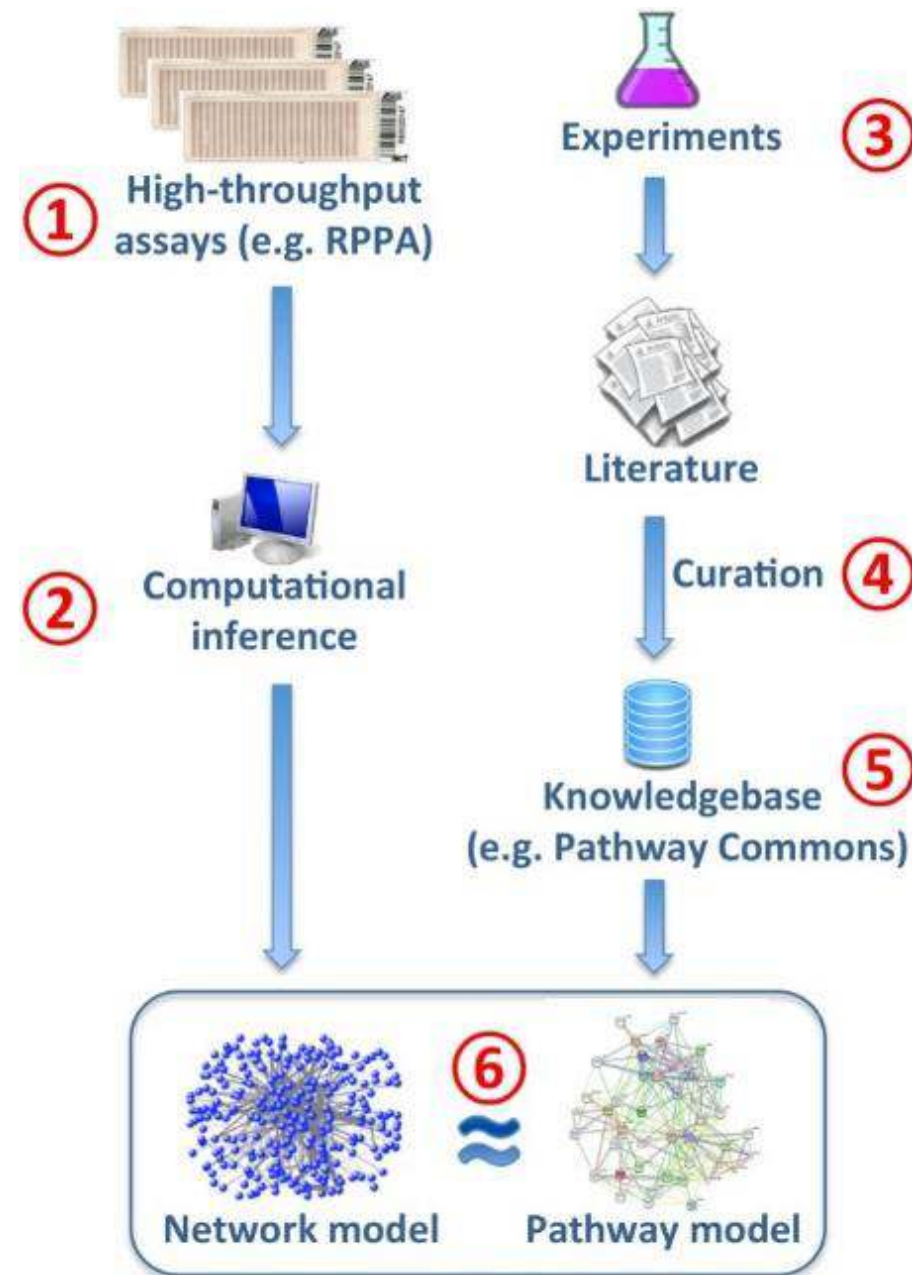
Воздействие загрязненного воздуха связано с неблагоприятными последствиями для здоровья, включая увеличение сердечно-легочной заболеваемости и смертности, а также увеличение риска рака легких. Был проведен полно-геномный анализ уровней экспрессии РНК крови 63 некурящих сотрудников парковочных терминалов в США. Экспрессия ряда генов изменилась при воздействии трех загрязнителей (твердых микрочастиц, элементарного углерода и органического углерода). В том числе экспрессия 63 гена, статистически значимо изменилась в случае двух из трех загрязнителей. Многие из этих генов были вовлечены в ишемическую болезнь сердца, хроническую обструктивную болезнь легких (ХОБЛ), рак легких и другие заболевания, связанные с загрязнением. Анализ генной сети (реконструированной с использованием GeneMANIA) позволил выявить 25 взаимосвязанных генов, которые были общими для всех трех загрязнений. Многие из них ассоциированы с раком, а также вовлечены в связывание металлов и регуляцию апоптоза. Эти результаты описывают молекулярную взаимосвязь между воздействием загрязненного воздуха и неблагоприятными последствиями для здоровья.



Сеть взаимосвязей между генами, экспрессия которых изменилась в ответ на воздействие загрязнения воздуха, реконструированная с использованием GeneMANIA. Анализ позволил выявить 25 взаимосвязанных генов, которые были общими для всех трех загрязнений.

## Мульти-метод для реконструкции сетей белок-белок взаимодействий, найденных в 11 видах раковых заболеваний человека.

Раковые клетки имеют особые характеристики уровней экспрессии белков и пост-трансляционных модификаций. Ресурс The Cancer Genome Atlas (TCGA) объединяет профили экспрессии белков 3467 образцов пациентов, страдающих от 11 типов опухолей. Эти протеомные данные могут быть использованы для компьютерной реконструкции сетей белок-белок взаимодействий (PPI) и последующего изучения общих характеристик различных типов опухолей. В этом исследовании было проведено сравнение эффективности 13 автоматических методов для реконструкции белок-белок с сетями белок-белок взаимодействий системы Pathway Commons, реконструированными вручную экспертами. Было показано, что ни один из методов не обладает наилучшими показателями во всех типах опухолей, но группа из шести методов, включая методы, основанные на корреляции, взаимной информации и регрессии, позволяют реконструировать сети, которые хорошо согласуются с сетями, построенными вручную экспертами.





Спасибо за внимание!