

# 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 (единий формат, единый способ графического и текстового представления, возможность выполнения поискового запроса ко всем данным). Преимущества интеграции 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.

**Ассоциативные генные сети** являются сложными, комплексными молекулярно-генетическими сетями, ассоциированными с какими либо биологическими процессами, фенотипическими признаками или заболеваниями. Типы базовых объектов, которыми оперируют ассоциативные генные сети, подразделяются на несколько классов, согласно их природе:

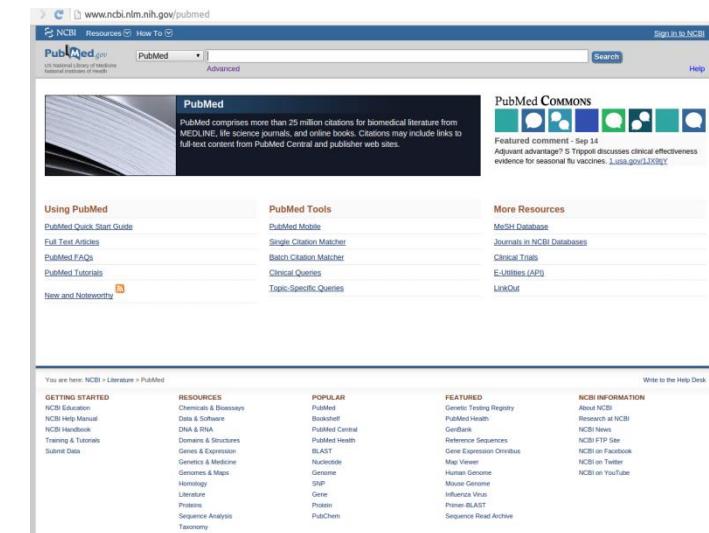
- (1) **молекулярно-генетические объекты**. К этому классу относятся такие типы объектов, как гены, РНК, белки, метаболиты, клеточные компоненты;
- (2) **биологические процессы и системы** (метаболические пути, пути передачи сигналов, транспортные пути и т.д.);
- (3) **фенотипические признаки**, поведенческие характеристики или функциональные состояния организма или генетических систем, включая заболевания;
- (4) **внутренние и внешние факторы**, воздействующие на систему (мутации, эпигенетический контроль, температура, давление, лекарства и другие химические соединения).

Таким образом, ассоциативные генные сети содержат существенно расширенный состав базовых объектов по сравнению с генными сетями, оперирующими только с молекулярно-генетическими объектами. Введение такого расширенного состава базовых объектов позволяет строить сети, в которых вершинами могут быть любые объекты из перечисленных классов. При этом, молекулярно-генетические объекты могут оказаться связанными с фенотипическими признаками или управляющими факторами. Такое свойство ассоциативных генных сетей позволяет проводить интеграцию разрозненных локальных генных сетей, которые напрямую не пересекаются между собой по молекулярно-генетическим объектам, но оказываются связанными на основании информации о их связи через какой либо расширенный объект, например фенотипический признак или заболевание.

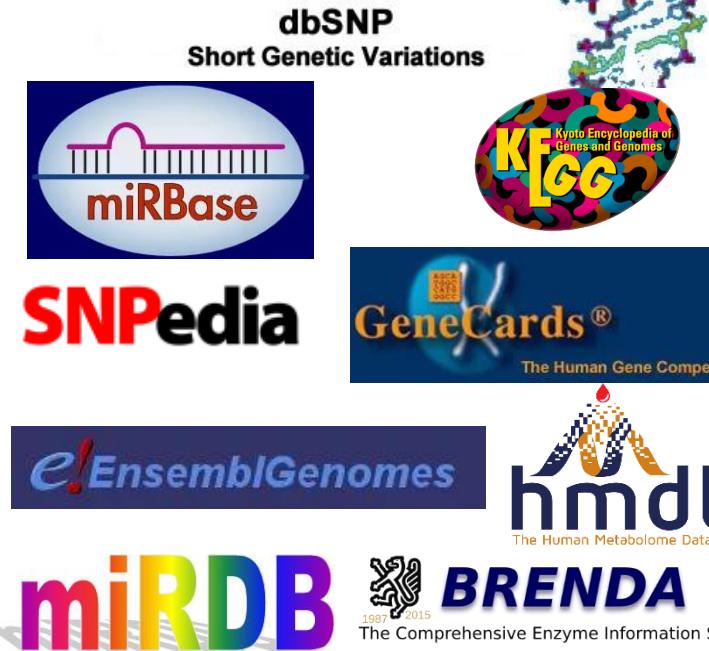
# Источники информации

- 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 Create RSS Create alert Advanced

Format: Summary Sort by: Most Recent Per page: 20 Send to: Filters: Manage Filters

Results by year

Download CSV

Related searches beta vulgaris antioxidant

Search results Items: 1 to 20 of 3864

<< First < Prev Page 1 of 194 Next > Last >

Seed-specific expression and analysis of recombinant anti-HER2 single-chain variable fragment (scFv-Fc) in Arabidopsis thaliana.

1. 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

Article types Clinical Trial Review Customize ... Text availability Abstract Free full text Full text PubMed Commons Reader comments Trending articles

Sign in to NCBI

PubMed Panax ginseng Create RSS Create alert Advanced

Format: Summary Sort by: Most Recent Per page: 20 Send to: Filters: Manage Filters

Results by year

Download CSV

Related searches panax ginseng review

Search results Items: 1 to 20 of 5999

<< First < Prev Page 1 of 300 Next > Last >

The Protective Effect of the Active Components of ERPC on Diabetic Peripheral Neuropathy in Rats.

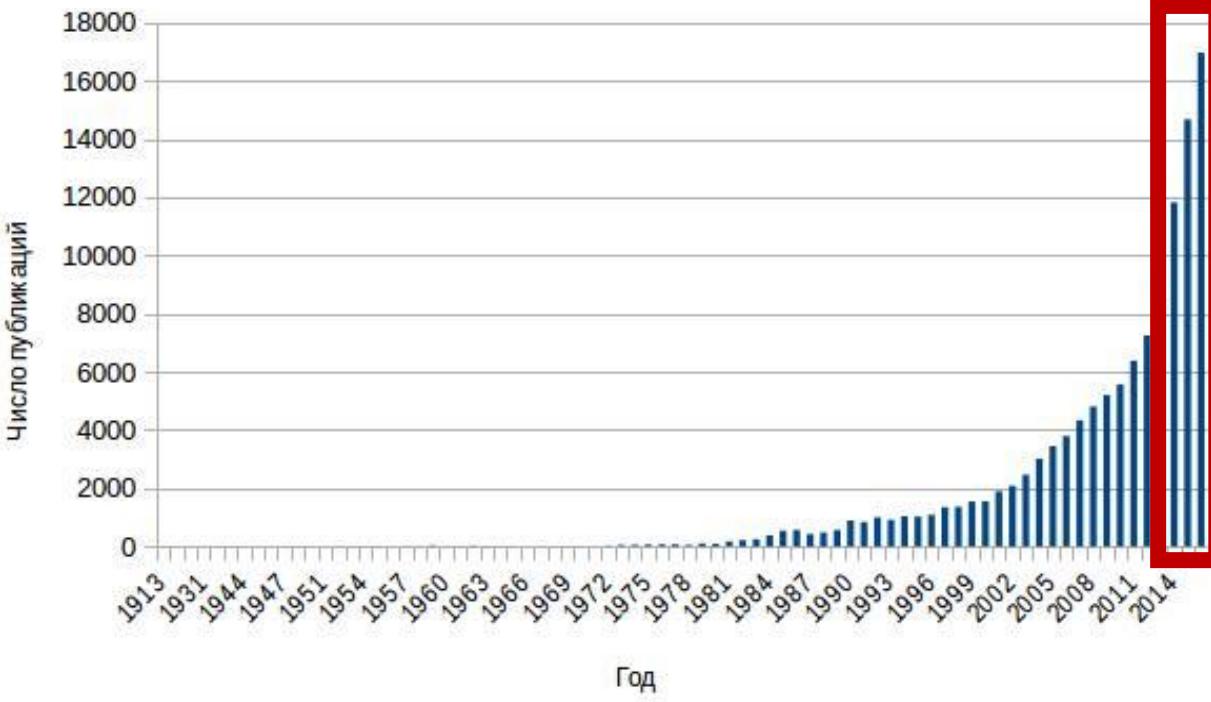
1. Hao GM, Liu YG, Wu Y, Xing W, Guo SZ, Wang Y, Wang ZL, Li C, Lv TT, Wang HL, Shi TJ, Wang W, Han J. J Ethnopharmacol. 2017 Mar 15; pii: S0378-8741(16)31590-2. doi: 10.1016/j.jep.2017.03.015. [Epub ahead of print]

PMID: 28315720 Similar articles

Article types Clinical Trial Review Customize ... Text availability Abstract Free full text Full text PubMed Commons Reader comments Trending articles

**Женьшень** (от кит. трад. 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 женьшена также используют в приготовлении пищи. Традиционная китайская медицина утверждает, что препараты женьшена 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.



Китайская традиционная медицина – альтернатива западным методам

11.12.2016 12 1900

VK +1 Twitter Facebook Поделиться 11

**И**нтерес к традиционной китайской медицине растет. Она выступает в качестве альтернативы западным методам в профилактике болезней и лечении хронических заболеваний. Об этом заявил Александр Чумбаков, заместитель главного врача одной из клиник традиционной китайской медицины в Москве.



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 1482 Редакция INFOPOL.RU

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 shows the HGNC Symbol Report for gene YY1. At the top, there's a navigation bar with links like Home, Downloads, Gene Families, Tools, Useful Links, About, Newsletters, Contact Us, Help, VGNC, and Request Symbol. A search bar is also present. The main content area is titled "Symbol Report: YY1". It displays various details about the gene:

- 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 this, there's a section for "External links" with tables for "HOMOLOGS" and "GENE RESOURCES". The "HOMOLOGS" table compares YY1 between Mus musculus (Yy1) and Rattus norvegicus (Yy1). The "GENE RESOURCES" table provides links to Entrez Gene, Ensembl, Vega, and UCSC databases. Further sections include "NUCLEOTIDE SEQUENCES", "PROTEIN RESOURCES", and "CLINICAL RESOURCES".

# 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

The screenshot shows the NCBI Gene search results page for the query 'yy1'. The search bar at the top contains 'yy1'. The results table lists several entries for the gene YY1 across different species:

Name/Gene ID	Description	Location	Aliases
YY1 ID: 7528	YY1 transcription factor [Homo sapiens (human)]	Chromosome 14, NC_000014.9 (100238765..100279034)	DELTA, INO80S, NF-E1, UCRBP, YIN-YANG-1
Yy1 ID: 22632	YY1 transcription factor [Mus musculus (house mouse)]	Chromosome 12, NC_000078.6 (108793269..108816632)	AW488674, NF-E1
Yy1 ID: 24919	YY1 transcription factor [Rattus norvegicus (Norway rat)]	Chromosome 6, NC_005105.4 (132702443..132726848)	NF-E1, NMP-1, NMP1, UCRBP
YY1 ID: 826093	zinc finger (C2H2 type) family protein [Arabidopsis thaliana (thale cress)]	Chromosome 4, NC_003075.7 (3764241..3766671, complement)	AT4G06634, AtYY1, Yin Yang 1
YY1 ID: 534353	YY1 transcription factor [Arabidopsis thaliana (thale cress)]	Chromosome 21, AC_000178.1 (30047744..30047854)	

Annotations on the left side of the results table include 'clear' and 'Show additional filters'. The right sidebar includes sections for 'Filters', 'Results by taxon', 'Find related data', 'Search details' (containing the query 'yy1[All Fields] AND alive[prop]'), and 'Recent activity'.

# NCBI Gene. Информационная карточка гена.

Идентификатор

Общая информация

Символ гена

Название гена

Ссылки на другие базы данных

Организм

Таксономическое положение

Синонимы

Описание гена

Ортологи

Более детальная информация

Расположение на хромосоме

Инtron/экзонная структура

Полиморфизмы

Литература и др.

The screenshot displays the NCBI Gene page for the YY1 transcription factor in *Homo sapiens*. The URL is <https://www.ncbi.nlm.nih.gov/gene/7528>. The page includes a sidebar with a table of contents and links to various databases and resources. The main content area shows the gene's official symbol (YY1), full name (YY1 transcription factor), primary source (HGNC:HGNC:12856), and taxonomic information (Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo). It also lists orthologs in mouse and all other species. Below this is a 'Genomic context' section showing the gene's location (14q32.2) and exon count (5), with a table of annotation details.

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
108	current	GRCh38.p7	14	NC_000014.9

# NCBI Nucleotide

- Доступна через интернет по адресу <https://www.ncbi.nlm.nih.gov/nuccore>
- 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 bar. The results page displays 3095 nucleotide sequences found across various species. The results are organized into a table with columns for sequence type, accession number, and GI number. Each result row includes links to GenBank, FASTA, and Graphics. On the left sidebar, there are filters for Species, Molecule types, Source databases, Genetic compartments, Sequence length, Release date, Revision date, and a 'Clear all' link. The right sidebar shows 'Results by taxon' with a list of top organisms and a 'Find related data' section with a database selection dropdown and a 'Search' button. The bottom sidebar shows a 'Recent activity' section with a search history for 'yy1'.

# NCBI Nucleotide. Карточка последовательности гена.

## Особенности первичной структуры последовательности

**Общая информация**

**Локус**

**Идентификатор**

**Организм и ТАКСОНОМИЯ** →

**Ссылки на литературу**

**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]

**Pathways for the YY1 gene**

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

Продолжение информационной карточки

**exon** 1384..1542  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
/inference="alignment:Splign:2.0.8"  
1543..3159  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
/inference="alignment:Splign:2.0.8"  
1729..1852  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
/standard\_name="RH1618"  
/db\_xref="UniSTS:[42115](#)"  
1954..2640  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
/standard\_name="YY1\_3892"  
/db\_xref="UniSTS:[462950](#)"  
2035..2173  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
/standard\_name="G06170"  
/db\_xref="UniSTS:[49520](#)"  
2690..2695  
/regulatory\_class="polyA\_signal\_sequence"  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
2714  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
2879..3086  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
/standard\_name="RH68412"  
/db\_xref="UniSTS:[51218](#)"  
3001..3100  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
/standard\_name="RH44921"  
/db\_xref="UniSTS:[36888](#)"  
3106..3111  
/regulatory\_class="polyA\_signal\_sequence"  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
3132  
/gene="YY1"  
/gene\_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"  
**STS**  
**STS**  
**STS**  
**regulatory**  
**polyA site**  
**STS**  
**STS**  
**regulatory**  
**polyA site**  
**ORIGIN**

1 agggcgaacg ggcgagtgcc agcgaggccg ggcgggctga gcccacgcg gaagtctcg  
61 gagggccggc cggagcacag tggtgcggcg gcggcggat ctgggctcg gttgaggagt  
121 tggatttgtt gtggaaaggag ggaggccgcg aggaggaaagg gggaaaggc ggcgcggcc  
181 ggaggccggg agggccgcg ccggaggccg gcgggttcgg cgaggcggg cgaggccggg  
241 agccgagacg acgacgcgc ggcgcgcgc gggcgcggc gcaccggc gaggaggccg  
301 gggaaaggcccc ggcgcggcc gcggccgcgc cccttcggc cccccc  
361 cggccgcctc ggcgccttc tcccttcggc ttcccttcggc acggccggcc gccttcggc  
421 cggccgcggcc gcagccgagg agccgaggcc gccgcggcc tgccggcggaa gcccctcggc

# 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 shows the GeneCards website interface for the YY1 gene. At the top, there's a navigation bar with links to GeneCardsSuite, GeneCards, MaloCards, LifeMap Discovery, PathCards, TGex, VarElect, GeneAnalytics, GeneALaCart, and GenesLikeMe. It also features a search bar and links for 'Keywords', 'Search Term', 'Advanced', 'My Genes', 'Log In / Sign Up', and social media icons for Twitter, LinkedIn, and Facebook.

The main content area displays the YY1 gene card. At the top left, it says "YY1 Gene (Protein Coding) ★" and "YY1 Transcription Factor". To the right, it shows "GCID: GC14P100238" and "GIFS: 64". Below this are sections for "Aliases", "Disorders", "Domains", "Drugs", "Expression", "Function", "Genomics", "Localization", "Orthologs", "Paralogs", "Pathways", "Products", "Proteins", "Publications", "Sources", "Summaries", "Transcripts", and "Variants".

Below the main title, there are four boxes containing logos for EMD Millipore, GenScript, ORIGENE, and Vigene Biosciences, each with associated product categories like Proteins & Enzymes, Antibodies, Peptides, etc.

The "Aliases for YY1 Gene" section lists several aliases: YY1 Transcription Factor, INO80 Complex Subunit S, Delta Transcription Factor, Yin And Yang 1 Protein, INO80S, NF-E1, YY-1, Transcriptional Repressor Protein YY1, Yin And Yang 1, YIN-YANG-1, Delta, and UCRBP.

The "External IDs for YY1 Gene" section provides HGNC, Entrez Gene, Ensembl, OMIM, and UniProtKB identifiers.

The "Previous GeneCards Identifiers for YY1 Gene" section lists various identifiers: GC14P098216, GC14P094520, GC14P098695, GC14P099774, GC14P100992, GC14P101002, GC14P080888, GC14P101020, GC14P101031, GC14P101051, GC14P101071, GC14P101112, GC14P101161, and GC14P101221.

The "Summaries for YY1 Gene" section includes an "Entrez Gene Summary for YY1 Gene" which describes YY1 as a zinc finger protein involved in repressing and activating diverse promoters through histone modifications.

# 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 паралогам
  - ✓ Биологическим процессам
  - ✓ Полиморфизмам
  - ✓ Экспрессии
  - ✓ Регуляции и др.

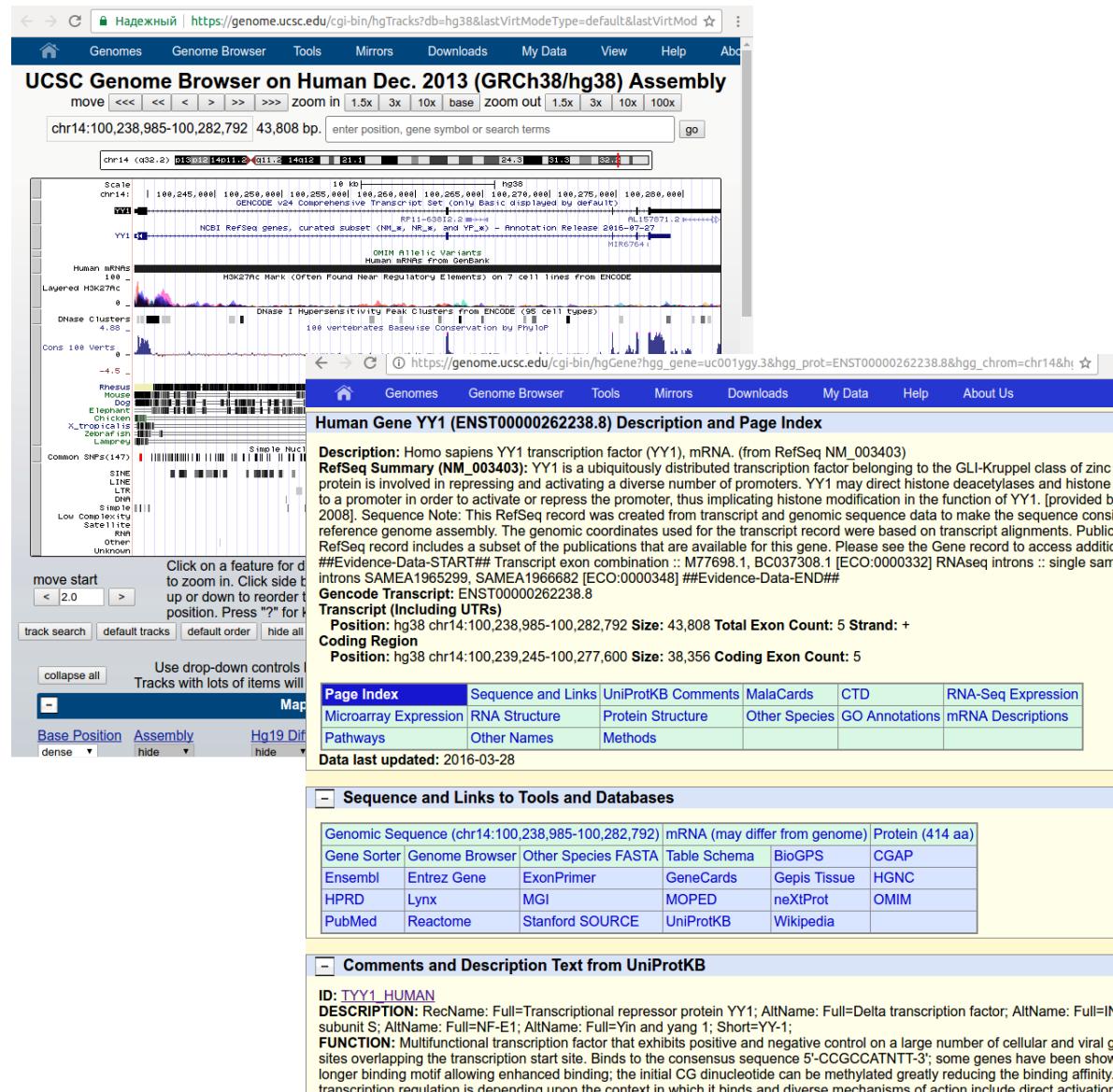
## Информационная карточка гена.

The screenshot displays the Ensembl gene summary page for YY1 in Human (GRCh38.p7). The URL is [www.ensembl.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000100811;r=14:100238298-100282792](http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000100811;r=14:100238298-100282792). The page header includes the Ensembl logo, navigation links (BLAST/BLAT, BioMart, Tools, Downloads, More), and a search bar. The main content area shows the gene identifier **Gene: YY1 ENSG00000100811**. To the left is a sidebar with "Gene-based displays" sections such as Summary, Sequence, Comparative Genomics, Ontologies, Phenotypes, Genetic Variation, ID History, and various configuration options (Configure this page, Custom tracks, Export data, Share this page, Bookmark this page). The main content area includes sections for Description (YY1 transcription factor), 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, Transcripts, and a Summary section with details like Name (YY1), CCDS, UniProtKB, RefSeq, Ensembl version (ENSG00000100811.11), Other assemblies (maps to 100,704,635-100,749,129 in GRCh37), Gene type (Known protein coding), Annotation method (includes automatic annotation from Ensembl and Havana manual curation), Alternative genes, and Annotation Attributes. At the bottom, there's a "Go to Region in Detail" link and a genomic track viewer showing the YY1 gene structure with multiple transcripts (YY1-005, YY1-002, YY1-004, RP11-638I2.6-001) and their respective features (retained intron, protein coding, lncRNA).

UCSC (University of California, Santa Cruz) Genome Browser

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

## Информационная карточка гена.



# 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 'Accession Display' page for experiment GSE14964. At the top, there's a header with the NCBI logo, the GEO logo, and links for GEO Publications, FAQ, MIAME, Email GEO, Not logged in, and Login. Below the header, there are search fields for Scope (Self), Format (HTML), Amount (Quick), and GEO accession (GSE14964). A button labeled 'GO' is also present. The main content area is titled 'Series GSE14964' and includes sections for Status, Title, Organism, Experiment type, Summary, Platforms, Samples, Relations, and Analyze with GEO2R. The 'Summary' section contains a detailed description of the experiment, mentioning YY1 and YY2 genes, shRNA constructs, and expression patterns. The 'Samples' section lists 13 samples, each with a unique identifier (e.g., GSM373614, GSM373615, etc.) and a brief description. The 'Relations' section links to BioProject PRJNA111841. The 'Analyze with GEO2R' section provides download options for SOFT, MINIIL, and TXT formats. At the bottom, there's a table for 'Supplementary file' showing GSE14964\_RAW.tar with a size of 251.4 Mb and a download link. Below the table, there are notes about raw and processed data files.

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.

The screenshot shows a web browser displaying the UCSC genome browser interface. The URL in the address bar is `genome.ucsc.edu/cgi-bin/hgGene?db=hg19&hgg_gene=uc003adj.3`. The page title is "Human Gene MN1 (uc003adj.3) Description and Page Index". A red box highlights the title "Human Gene MN1 (uc003adj.3) Description and Page Index". Below the title, there is a "Description" section stating "Homo sapiens meningioma (disrupted in balanced translocation) 1 (MN1), mRNA." and a "RefSeq Summary (NM\_002430)" section providing details about the gene's structure and publications. Further down, there are sections for "Transcript (Including UTRs)", "Position", "Coding Region", and "GO Annotations". At the bottom, there is a table with links to various databases and a note about the last update date.

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

Page Index	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 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: "X002000.2\_XY066700.1[EC00-00000221] RNA intron "

single sample

Transcript (1)  
Position: 1  
Coding Region Position: 1

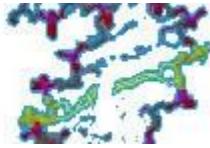
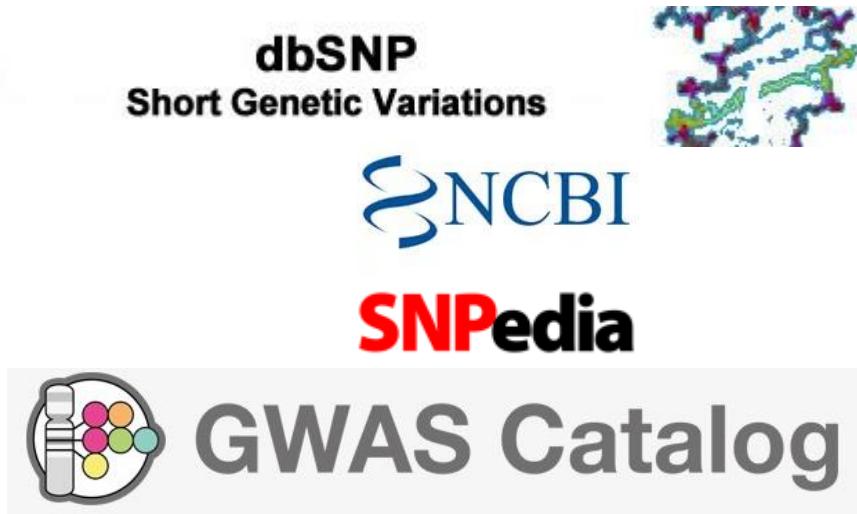
Known Genes

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
- 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 Reference SNP (refSNP) Cluster Report for rs61992955. The top navigation bar includes links for dbVar, ClinVar, GaP, PubMed, Nucleotide, and Protein. A search bar allows searching for small variations in dbSNP or large structural variations in dbVar. The main content area is titled "Reference SNP (refSNP) Cluster Report: rs61992955".  
Key details shown:

- RefSNP:** Organism: human (*Homo sapiens*); Molecule Type: Genomic; Created/Updated in build: 129/149; Map to Genome Build: 108/Weight 1; Validation Status:
- Allele:** Variation Class: SNV: single nucleotide; RefSNP Alleles: C/G (FWD); Allele Origin: ; Ancestral Allele: C; Variation Viewer: VarView; Clinical Significance: NA; MAF/MinorAlleleCount: C=0.0300/150 (
- Integrated Maps:** Shows GRCh38.p7 and GRCh37.p13 assemblies with Chr 14 and positions 100274488 and 100740825 respectively, with links to NT\_026437.13 and NT\_026437.12.
- GeneView:** Details the YY1 transcription factor annotation across contigs, showing assembly GRCh38.p7, annotation release 108, and position 14. It includes a "View more variation on this gene" link and clinical source options.
- Primary Assembly Mapping:** Lists GRCh38.p7, Fwd, Chr 14, and position 100274488.
- RefSeqGene Mapping:** Lists RefSeqGene NG\_046908.1, Gene (ID) YY1 (7528), and SNP to Refs 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 разветвленные 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.

NM\_003403.4(YY1):c.1115C>G (p.Thr372Arg)

Variation ID: 91950  
Review status: (0/4) no assertion criteria provided

Interpretation

Clinical significance: Uncertain significance  
Number of submission(s): 1  
See supporting ClinVar records

Allele(s)

Allele ID: 97428  
Variant type: single nucleotide variant  
Cytogenetic location: 14q32.2  
Genomic location:

- Chr14: 100277470 (on Assembly GRCh38)
- Chr14: 100743807 (on Assembly GRCh37)

  
Protein change: T372R  
HGVS:

- NG\_046908.1:g.43706C>G
- NM\_003403.4:c.1115C>G
- NP\_003394.1:p.Thr372Arg

Links:

- UniProtKB: P25490#VAR\_074172
- dbSNP: 386834266

NCBI 1000 Genomes Browser: rs386834266

Molecular consequence: NM\_003403.4:c.1115C>G: missense variant [Sequence Ontology SO:0001583]

1 Affected gene  
YY1 transcription factor (YY1) [Gene OMIM - Variation Viewer]  
Search ClinVar for variants within YY1  
Search ClinVar for variants including YY1

Variant frequency in dbGaP  
NM\_003403.4(YY1):c.1115C>G (p.Thr372Arg)  
GRCh37 Chr14:100743807

Called variants	Potential variants
Sample count	no data 0 of 4090

Called variants are samples submitted to db that have the variant allele. Potential variants SRA runs that display the allele in at least 30% of the reads covering the position, and have 10x more passing reads covering the position.

Browser views  
RefSeqGene  
Variation Viewer [GRCh38 - GRCh37]  
UCSC [GRCh38/hg38 - GRCh37/hg19]

Related information  
dbSNP  
Gene  
MedGen  
Related genes (specific)

# 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.

The screenshot displays the SNPedia article for the SNP rs2494262. At the top, there's a navigation bar with links for 'Статья' (Article), 'Обсуждение' (Discussion), 'Редактировать с помощью формы' (Edit with form), 'Править' (Edit), and 'История' (History). Below the navigation is a message: 'Have questions? Visit <https://www.reddit.com/r/SNPedia>'. The main content area is titled 'rs2494262'. It contains several text snippets with PMID links, such as 'FepsilonRIalpha gene -18483A>C polymorphism affects transcriptional activity through YY1 binding' and 'Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus'. To the right of the text is a sidebar with a table of allele frequency data:

Orientation	plus
Stabilized	plus
Make rs2494262(A;A)	
Make rs2494262(A;C)	
Make rs2494262(C;C)	

Below the table, more detailed information is provided:

Reference	GRCh38
38.1/141	
Chromosome	1
Position	159283882
Gene	FCER1A
GWAS Ctlg	rs2494262
GMAF	0.4601
Max	
Magnitude	

At the bottom, there's a population frequency bar chart. The legend indicates three alleles: (A;A) in brown, (A;C) in green, and (C;C) in blue. The x-axis represents frequency from 0 to 100. The y-axis lists populations: CEU, HCB, JPT, YRI, ASW, CHB, CHD, GIH, LWK, MEX, MKK, TSI, and AVG. The chart shows that for most populations, the (A;A) allele is dominant, while (C;C) is very low or absent.

# 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.**

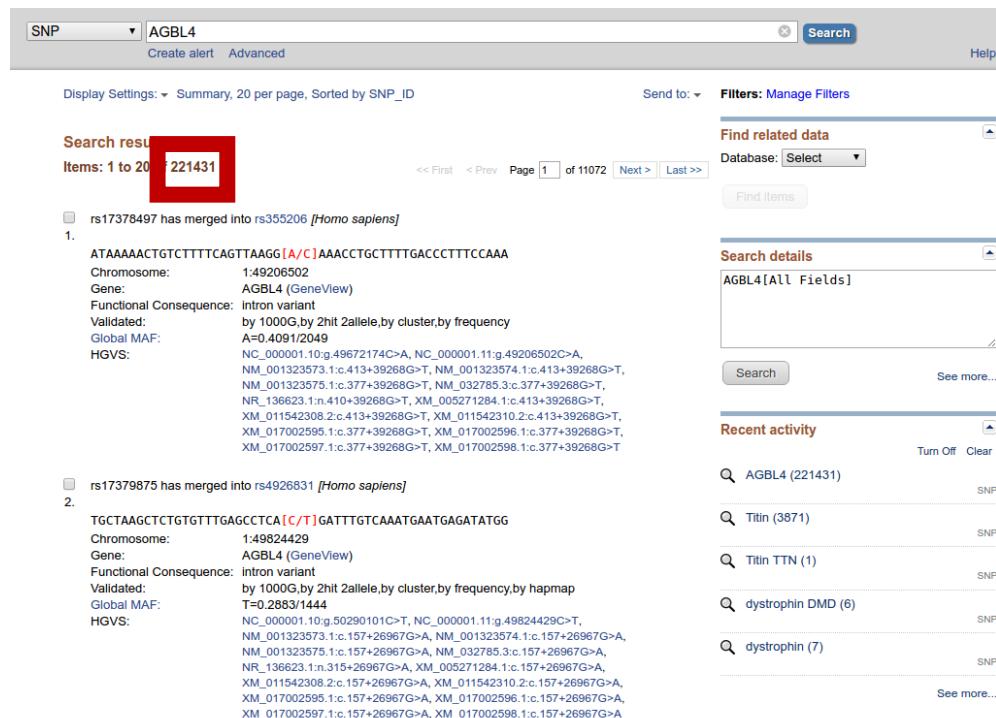
The screenshot shows the GWAS Catalog search results for the SNP rs2766692. The top navigation bar includes links for Home, Search, Diagram, Download, Documentation, About, EMBL-EBI, and NIH. The main content area features a circular graphic representing genetic data. Below it, the title "GWAS Catalog" is displayed, followed by the subtitle "The NHGRI-EBI Catalog of published genome-wide association studies". A search bar contains the identifier "rs2766692". A list of study associations is shown, with one study highlighted: "Kang SJ (PMID: 22554406)" from "2012-05-03" in "Genes Brain Behav". The study title is "Family-based genome-wide association study of frontal θ oscillations identifies potassium channel gene KCNJ6." The search interface also includes filters for "Studies", "Associations", and "Catalog traits", and options to "Refine search results" and "Download association results".

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 AGBL4 Search Help

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

Send to: Filters: Manage Filters

Find related data Database: Select

Find items

Search results: Items: 1 to 20 of 221431

rs17378497 has merged into rs355206 [Homo sapiens]

1. ATAAACTGTCTTCAAGTAAAG [A/C] AACACTGCTTTGACCCCTTCAAA

Chromosome: 1:49206502  
Gene: AGBL4 (GeneView)  
Functional Consequence: intron variant  
Validated: by 1000G, by 2hit, 2 allele, 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\_001323575.3: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

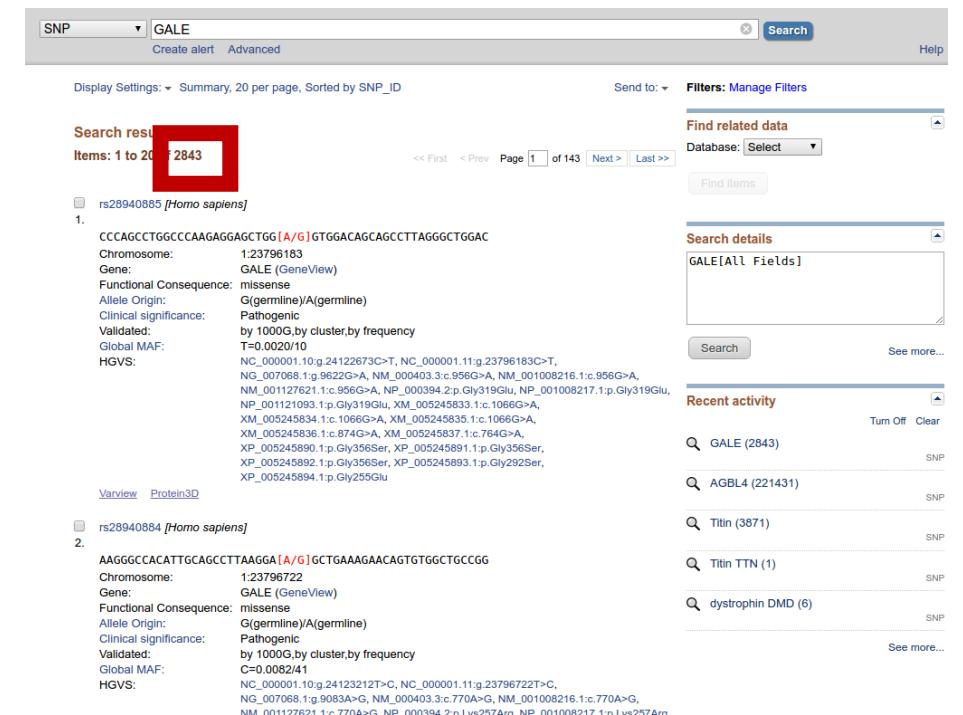
rs17379875 has merged into rs4926831 [Homo sapiens]

2. TGCTAACCTGTGTTGAGCCTCA [C/T] GATTGTCAAATGAATGAGATATGG

Chromosome: 1:49824429  
Gene: AGBL4 (GeneView)  
Functional Consequence: intron variant  
Validated: by 1000G, by 2hit, 2 allele, 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

### 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 GALE Search Help

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

Send to: Filters: Manage Filters

Find related data Database: Select

Find items

Search results: Items: 1 to 20 of 2843

rs28940885 [Homo sapiens]

1. CCGAGCTGGCCAAGAGGAGCTGG [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, NM\_000394.1: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\_005245834.1:p.Gly319Glu, 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

2. rs28940884 [Homo sapiens]

AAGGGCACCATGTGAGCCTTAAGGA [A/G] GCTGAAAGAACGTGTGGCTGGCG

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.24123212C>T, NC\_000001.11:g.23796722C>T, 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:n.1 vs 257A>n.1 vs 257A

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: chr1:48,998,527-50,489,62 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

This screenshot shows the dbSNP page for the AGBL4 gene. At the top, it displays the gene name, chromosome position (chr1:48,998,527-50,489,62), size (1,491,100), and exon count (14). Below this, there are sections for 'Coding Region' and 'Position' with a different set of coordinates (chr1:48,999,845-50,489,468) and sizes (1,489,624). The main content area shows search results for SNPs, with the first result being rs17378497. The results table includes columns for ID, Chromosome, Position, Allele, and HGVS. There are also filters for 'Find related data' and 'Search details'.

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

Description: UDP-galactose-4-epimerase  
Transcript (Including UTRs):  
Position: chr1:23,994,676-23,999,88 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

This screenshot shows the dbSNP page for the GALE gene. It displays the gene name, chromosome position (chr1:23,994,676-23,999,88), size (5,206), and exon count (12). Below this, there is a section for 'Coding Region' with a different set of coordinates (chr1:23,995,026-23,998,084) and sizes (3,059). The main content area shows search results for SNPs, with the first result being rs28940885. The results table includes columns for ID, Chromosome, Position, Allele, and HGVS. There are also filters for 'Find related data' and 'Search details'.

# 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.

The screenshot shows the UniProtKB entry page for P25490 (TYY1\_HUMAN). The top navigation bar includes links for BLAST, Align, Retrieve/ID mapping, Peptide search, Advanced search, and a search bar. The main title is "UniProtKB - P25490 (TYY1\_HUMAN)". On the left, there is a sidebar titled "Display" with tabs for Entry, Publications, Feature viewer, and Feature table. The "Entry" tab is selected, showing the following details:

- Protein: Transcriptional repressor protein YY1
- Gene: YY1
- Organism: Homo sapiens (Human)
- Status: Reviewed - Annotation score: 5/5 - Experimental evidence at protein level

The main content area is titled "Function" and describes the protein as a 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. It 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

# 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 ААН65366.1
- 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 information card for the YY1 transcription factor from Homo sapiens. The card includes the following details:

**GenBank:** AAH65366.1  
**Identical Proteins:** FASTA, Graphics  
**Go to:** AAH65366.1  
**Region:** C2H2 Zn finger (385..407)  
**Site:** Zn binding site (ion binding) (385, 390, 403, 407)  
**CDS:** YY1 (1..414)  
**ORIGIN:** Amino acid sequence (partial):  
masgdlyia tdgsempaei velheievt ipvetiettv vgeeeeeeddd dedggggdhg  
61 ggghghagh hhhhhhhhh ppmialqlpv dddptqvhhh qevilvqtre evvgggdsdg  
121 lraedgfedq iliprvapag gdddyieqtl vtvaagksq gggssssggg rvkkgggkks  
181 gkksylsgga gaaggggadp gnkkweqkqv qiktlegefs vtmwssdekk didhetvvve  
241 qiqengsppd yseymtgkkl pppgipgidl sdpkqlaefa rmkprkiked daprtiacph  
301 kgctkmfrdn samrkhllth gprvhvcaec gkafvesskl krhqlvhtge kpfqctfegc  
361 gkrfsldfnl rthvrihtgd rpyvcfdgc nkkfaqstnl kshilthaka knng  
//

**Protein 3D Structure:** Crystal Structure O Mbt1 YY1 Complex PDB: 4C5I Source: Homo sapiens

**LinkOut to external resources:** A selection of literature about the proteins [GoPubMed Proteins]

**MODBASE:** Database of Comparative P [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] [Print]

(yy1) AND "Homo sapiens"[orgn] [Print]

# 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.

The screenshot shows the RCSB PDB website interface for protein 1DGB. At the top, the header includes the RCSB PDB logo, navigation links (Deposit, Search, Visualize, Analyze, Download, Learn, More), and a 'MyPDB Login' button. Below the header, the PDB logo is displayed with the text 'An Information Portal to 127823 Biological Macromolecular Structures'. A search bar allows users to search by PDB ID, author, macromolecule, sequence, or ligand. Below the search bar are links for 'Advanced Search', 'Browse by Annotations', 'Search History (7)', and 'Previous Results (6)'. The main content area features a large 3D ribbon diagram of the Human Erythrocyte Catalase protein. To the left of the structure is a 'Structure Summary' panel containing links to 'View in 3D' (NGL or JSmol or PV), 'Standalone Viewers' (Simple Viewer, Protein Workshop, Ligand Explorer, Kiosk Viewer), 'Protein Symmetry' (Dihedral - D2), and 'Protein Stoichiometry' (Homo 4-mer - A4). The right side of the card displays detailed information about the protein: ID 1DGB, Classification OXIDOREDUCTASE, DOI 10.22110/pdb1dgb/pdb, 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, and Structural Biology Knowledgebase annotations (SBKB.org). Below this is an 'Experimental Data Snapshot' section with metrics like Method X-RAY DIFFRACTION, Resolution 2.2 Å, R-Value Free 0.227, and R-Value Work 0.172. To the right is a 'wwPDB Validation' section with a chart showing Clashscore, Ramachandran outliers, Sidechain outliers, and RSRZ outliers relative to all X-ray structures and similar resolution structures. At the bottom, a 'Literature' section lists the reference: 'Active and inhibited human catalase structures: ligand and NADPH binding and catalytic mechanism.' by Putnam, C.D., Arvai, A.S., Bourne, Y., Tainer, J.A. (2000) J.Mol.Biol. 296: 295-309, with a PubMed link 10656833.

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?

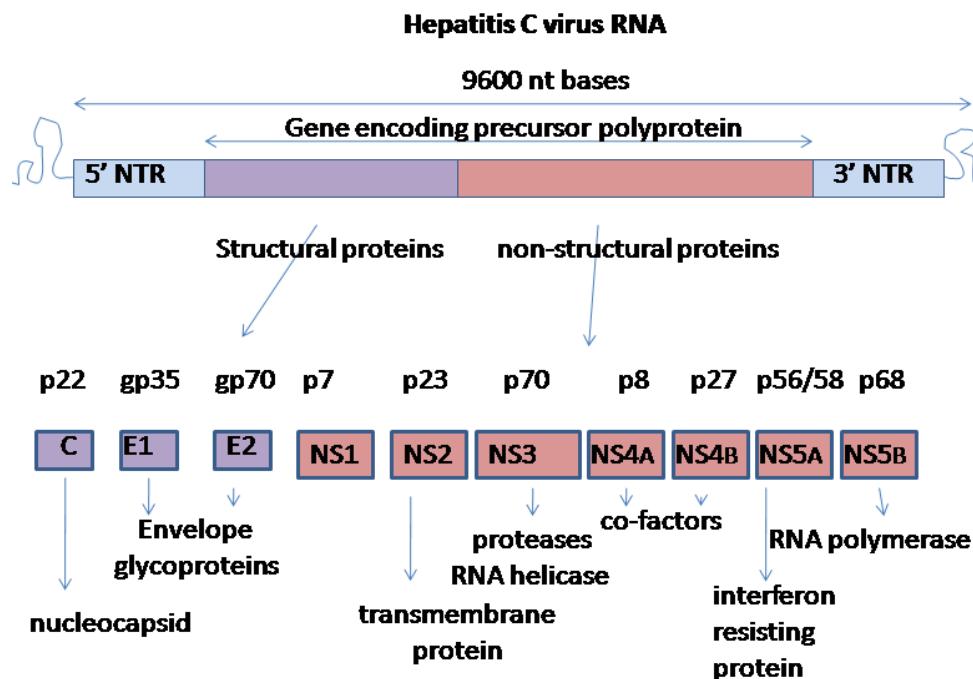


## Информационная карточка UniProtKB - P27958 (POLG\_HCVH)

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



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      p23
- name:      Serine protease NS3 (EC:3.4.21.98, EC:3.6.1.15, EC:3.6.4.13)

Alternative name(s):

- Hepacivirus
- NS3P
- p70

• Non-structural protein 4A

- Short      NS4A
- name:      Alternative name(s):
- p8

• Non-structural protein 4B

- Short      NS4B
- name:      Alternative name(s):
- p27

• Non-structural protein 5A

- Short      NS5A
- name:      Alternative name(s):
- p56

• RNA-directed RNA polymerase (EC:2.7.7.48)

Alternative name(s):

- NS5B
- p68

Organism<sup>i</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.

The screenshot shows the miRBase website interface for the miRNA mmu-mir-302b. At the top, there's a navigation bar with links for Home, Search, Browse, Help, Download, Blog, Submit, and the specific entry for mmu-mir-302b. The main content area has a dark blue header "Stem-loop sequence mmu-mir-302b". Below it, there's a table with columns for Accession (MI0003716), Symbol (MGI:Mir302b), Description (Mus musculus miR-302b stem-loop), and Gene family (MIPF0000071; mir-302). A large box displays the stem-loop sequence with colored nucleotides (green, red, blue) and positions. A "Get sequence" button is below the sequence. To the right, there's a bar chart titled "1633 reads, 84.5 reads per million, 35 experiments" with a pink gradient. Below the chart is a sequence logo. A feedback section asks "Feedback: Do you believe this miRNA is real?" with options "Yes (+4)", "No (-0)", and "Leave comment". A detailed comments section explains the experimental verification of the mature sequence. At the bottom, genome context information is provided, including coordinates (chr3: 127545228-127545301 [+]), overlapping transcripts (antisense ENSMUST0000029588; Larp7-201; intron 8), and other details.

# miRDB

- Доступна через интернет по адресу <http://www.mirdb.org/miRDB/index.html>
- Содержит информацию по структуре и мишениям 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 именования (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).

## Информационная карточка микроРНК.



MicroRNA and Target Gene Description:			
<b>miRNA Name</b>	hsa-miR-302b-3p	<b>miRNA Sequence</b>	UAAGUGCUUCCAUGUUUUAGUAG
<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>	55432	<b>GenBank Accession</b>	NM_018566
<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 TCCAGAACGC TCTGGGTTTT
 6 CCAATAAGCT ATGGTAACCC TAAAGAACAA AGGATAACAAT GCTTGAACCA TCCTTTAAC
121 TTAAAACCAC TAAGACACTG AAATTCTTG TTAAGATTAA ATTAGTGTG CAAGTTTACA
181 GATGTGTGTC TACAGTGGTA AACTGTACAT ACATGCCCTC TTCTGCTGGA GTGACAGAAT
241 AGGTGTATCTT TCCCACCTAC TGACACTGAC CTGAAGGTTG AGATTGACTA TTATTAATCA
```



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="#">LAT52</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 subunit

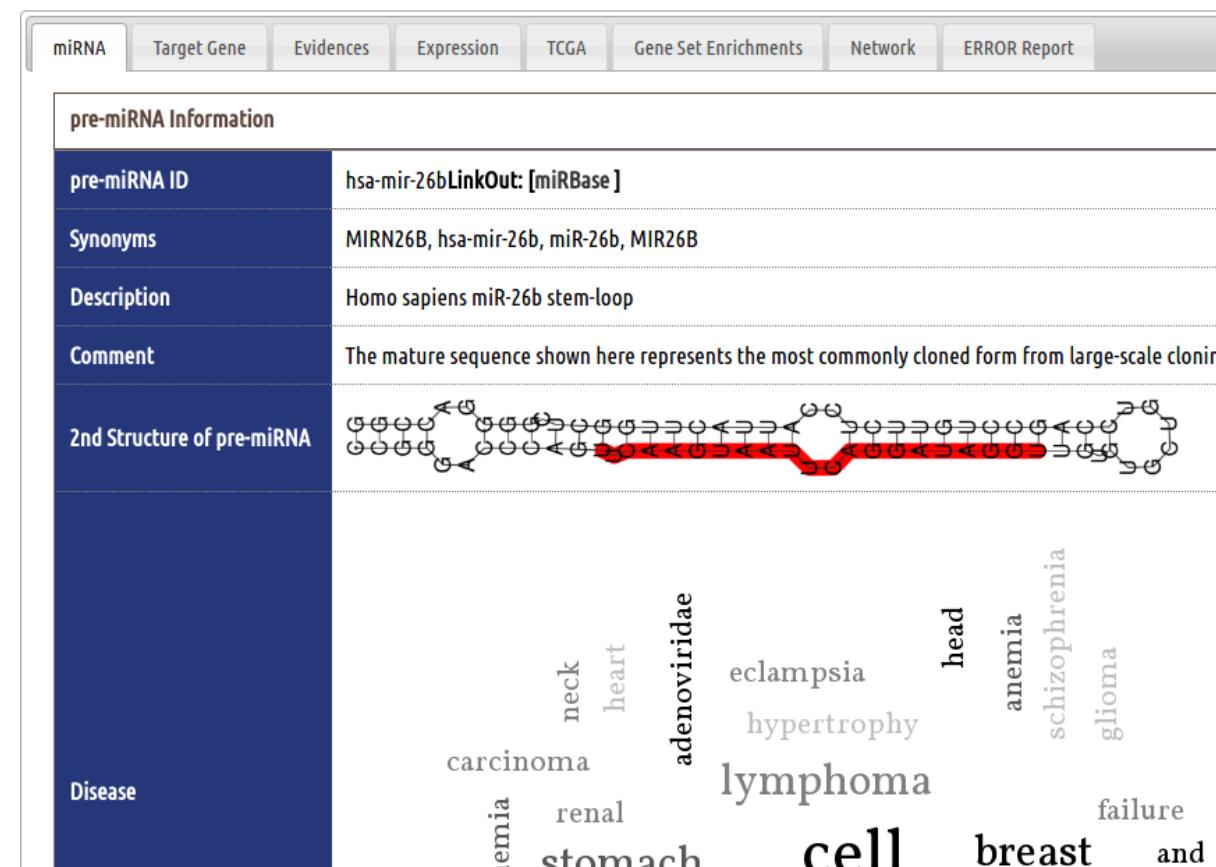
# miRTarBase

- 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-5р имеет идентификатор MIRT029499.
- Содержит информацию по
  - ✓ Мишеням
  - ✓ Последовательности шпильки
  - ✓ Синонимам
  - ✓ Ассоциированным заболеваниям
  - ✓ Ссылки на публикации и др.

Информационная карточка микроРНК.

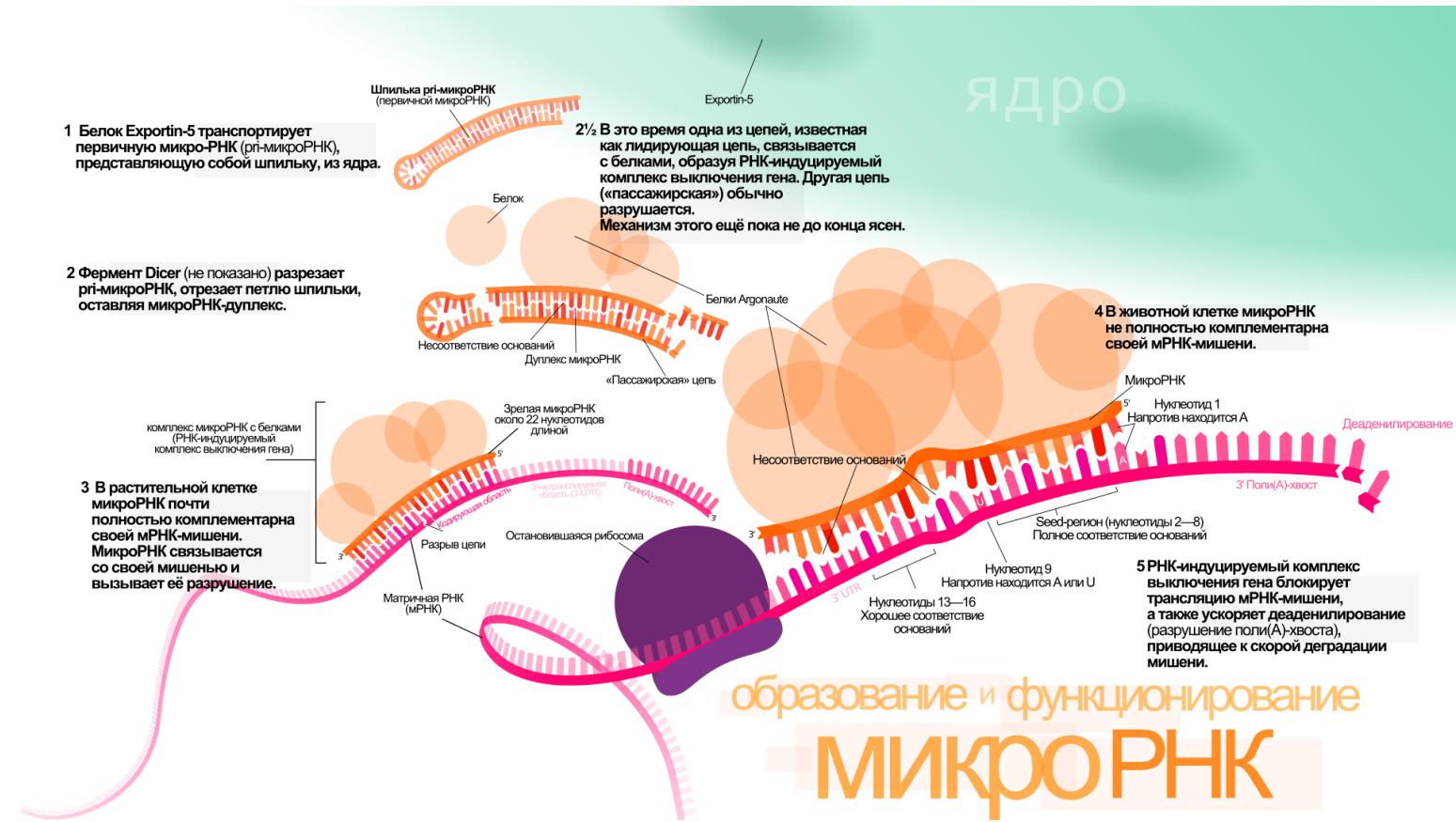
## miRTarBase

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



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.

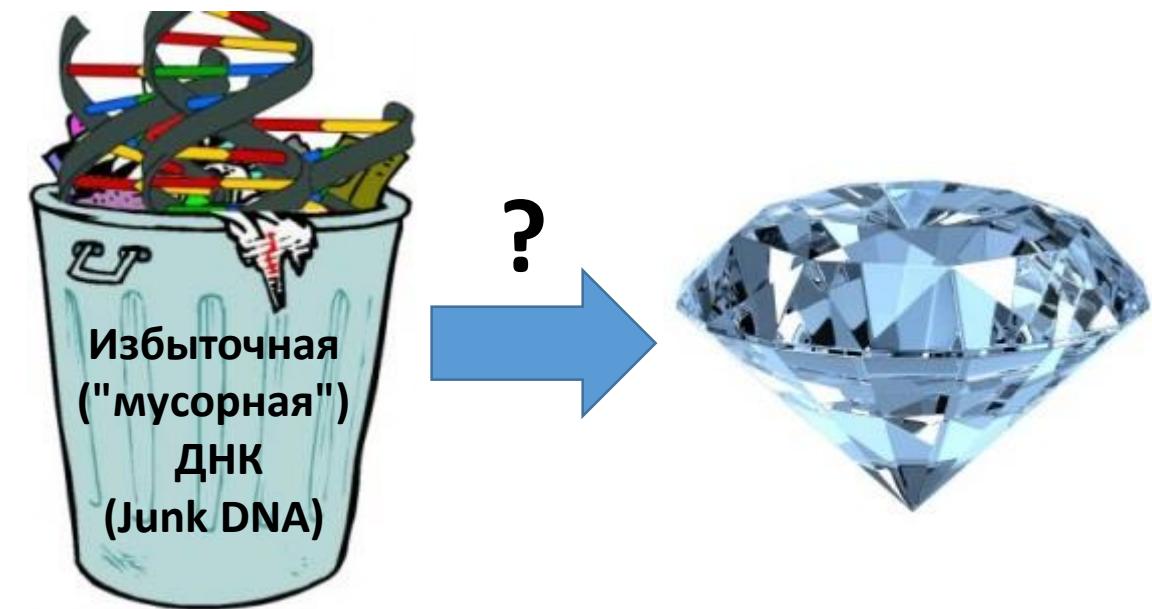
А какие еще классы некодирующих РНК Вы можете назвать?



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. vlncRNA – very long non-coding RNA



# Базы данных, содержащие информацию по метаболитам.



ChEBI



- 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 for the compound with ID CHEBI:15347, which is acetone. The page includes a navigation bar with links to Home, Advanced Search, Browse, Documentation, Download, Tools, About ChEBI, Preferences, Submit, and Contact us. A search bar at the top right contains the identifier 'CHEBI:15347'. Below the search bar, there are tabs for Main, ChEBI Ontology, Automatic Xrefs, Reactions, Pathways, and Models. The Main tab is selected. The main content area displays the chemical structure of acetone (CH3C(=O)CH3), its ChEBI Name ('acetone'), ChEBI ID ('CHEBI:15347'), and its definition ('A methyl ketone that consists of propane bearing an oxo group at C2.'), which is highlighted in blue. It also shows a rating of two stars ('★★') and a note that it has been manually annotated by the ChEBI Team. Other sections include Secondary ChEBI IDs (CHEBI:40571, CHEBI:2398, CHEBI:13708, CHEBI:22182), Supplier Information (ZINC00895111, eMolecules:474422), and a 'Download Molfile' button. At the bottom, there are links to 'Find compounds which contain this structure', 'Find compounds which resemble this structure', and 'Take structure to the Advanced Search'. A 'more structures >>' link is also present. The bottom of the page includes a 'Wikipedia' link and a 'License' link.

# 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.

The screenshot displays the 'Compound Summary for CID 180' page for Acetone. At the top, the PubChem logo and search bar are visible. Below the header, there are tabs for Structure, Vendors, Pharmacology, Literature, Patents, and Bioactivities. Key data points include:

- PubChem CID:** 180
- Chemical Names:** Acetone; 2-propanone; Propanone; Dimethyl ketone; Methyl ketone; 67-64-1
- 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:** CSCPPACGZOOCGX-UHFFFAOYSA-N
- Substance Registry:** FDA UNII
- Safety Summary:** Laboratory Chemical Safety Summary (LCSS)

Below this, a detailed description of Acetone is provided, mentioning its use as a solvent and antiseptic, and noting it is one of the ketone bodies produced during ketoacidosis. A link to the Pharmacology section from NCIt is also present. The bottom of the page contains a detailed pharmacological summary of Acetone's role in metabolism and its effects on the body.

# 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 ( $\beta$ -hydroxybutyrate, 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.

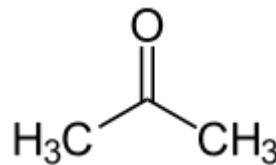
Является естественным метаболитом, производимым организмами млекопитающих. Некоторое количество вещества выводится с выдыхаемым воздухом и выделениями кожи, некоторое — с мочой. В крови человека в норме содержится 1-2 мг/100 мл этого вещества, в суточном количестве мочи — 0,01-0,03 г. При нарушениях 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 различном производстве 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



# Базы данных, содержащие информацию по заболеваниям.

OMIM®



DISEASE  
ONTOLOGY

- 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.

The screenshot shows the OMIM website interface. At the top, there is a navigation bar with links for About, Statistics, Downloads, Contact Us, MIMmatch, Donate, Help, and a question mark icon. Below the navigation bar is a search bar containing the text "diabetes". To the right of the search bar are a magnifying glass icon and an "Options" dropdown menu. On the left side of the main content area, there is a sidebar titled "External Links" which includes links for Protein, Clinical Resources (Clinical Trials, Genetic Alliance, Genetics Home Reference GTR), Animal Models, and Cell Lines. The main content area displays the following information:

- # 125853**
- 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**: A table showing associations between chromosomes, phenotypes, MIM numbers, inheritance patterns, mapping keys, genes/loci, and MIM numbers.

Location	Phenotype	MIM number	Inheritance	Phenotype mapping key	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, a postsurgical hypoinsulinaemia - E89.1.

ICD-10 Version:2010

Search diabetes [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	.1 With ketoacidosis	.2+ With renal complications	.3+ With ophthalmic complications	.4+ With neurological complications
Diabetic: <ul style="list-style-type: none"><li>coma with or without ketoacidosis</li><li>hyperosmolar coma</li><li>hypoglycaemic coma</li></ul> Hyperglycaemic coma NOS	Diabetic: <ul style="list-style-type: none"><li>acidosis</li><li>ketoacidosis</li></ul> without mention of coma	Diabetic nephropathy (N08.3*) Intracapillary glomerulonephrosis (N08.3*) Kimmelstiel-Wilson syndrome (N08.3*)	Diabetic: <ul style="list-style-type: none"><li>cataract (H28.0*)</li><li>retinopathy (H36.0*)</li></ul>	Diabetic: <ul style="list-style-type: none"><li>amyotrophy (G73.0*)</li><li>autonomic neuropathy (G99.0*)</li><li>mononeuropathy (G59.0*)</li><li>polyneuropathy (G63.2*)<ul style="list-style-type: none"><li>autonomic (G99.0*)</li></ul></li></ul>

# 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 a web browser displaying the Disease Ontology website at [disease-ontology.org](http://disease-ontology.org). The page title is "DISEASE ONTOLOGY". A search bar at the top contains the term "diabetes". The main content area is titled "Metadata" and displays the following information for "Diabetes Mellitus":

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. <a href="http://www.ncbi.nlm.nih.gov/pubmed/?term=9686693">http://www.ncbi.nlm.nih.gov/pubmed/?term=9686693</a> , <a href="http://www.who.int/diabetes/action_online/basics/en/">http://www.who.int/diabetes/action_online/basics/en/</a> , <a href="https://en.wikipedia.org/wiki/Diabetes_mellitus">https://en.wikipedia.org/wiki/Diabetes_mellitus</a>
Xrefs	<a href="#">ICD10CM:E08-E13</a> <a href="#">ICD10CM:E11</a> <a href="#">ICD9CM:250</a> <a href="#">MSH:D003920</a> <a href="#">NCI:C2985</a> <a href="#">SNOMEDCT_US_2016_03_01:154671004</a> <a href="#">SNOMEDCT_US_2016_03_01:191044006</a> <a href="#">SNOMEDCT_US_2016_03_01:267467004</a> <a href="#">SNOMEDCT_US_2016_03_01:73211009</a> <a href="#">UMLS_CUI:C0011849</a>
Relationships	is_a <a href="#">glucose metabolism disease</a>

On the left side, there is a navigation tree under the heading "Navigation" with the following structure:

- disease
  - disease by infectious agent
  - disease of anatomical entity
  - disease of cellular proliferation
  - disease of mental health
  - disease of metabolism
    - acquired metabolic disease
      - amyloidosis
      - carbohydrate metabolism disease
        - glucose metabolism disease
          - diabetes mellitus
          - hyperglycemia
          - hyperinsulinism
          - hypoglycemia
          - triosephosphate isomerase
        - carotenemia
        - hyperprolactinemia
        - hyperuricemia
        - lactic acidosis
        - mineral metabolism disease
        - nutrition disease
        - inherited metabolic disorder
        - genetic disease
        - physical disorder
        - syndrome

Оцените, сколько в настоящее время известно заболеваний?

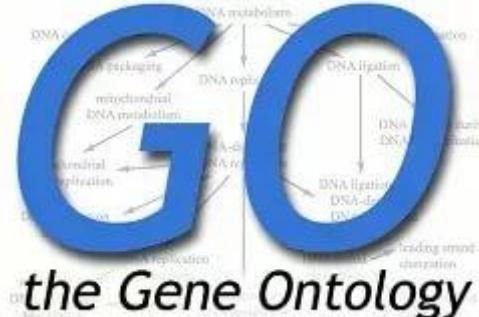


Оцените, сколько в настоящее время известно заболеваний?

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.

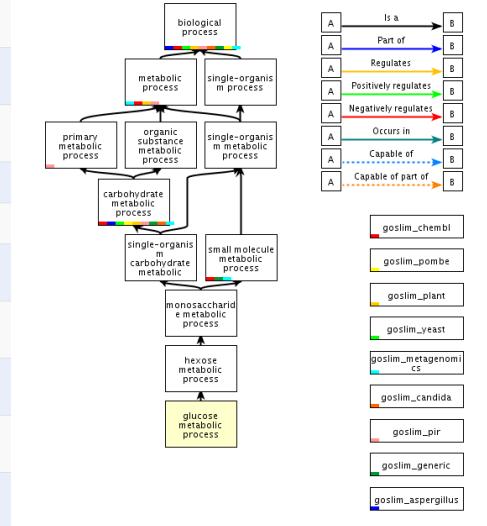
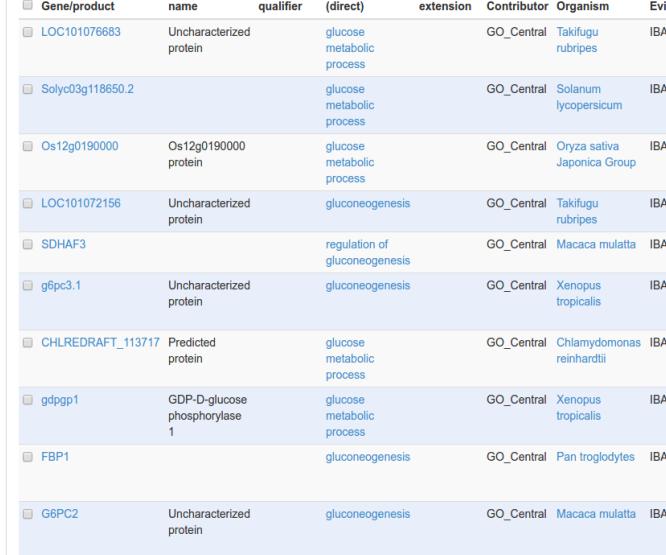
glucose metabolic process

Term Information 

Data health 

Accession	GO:0006006
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 <a href="#">history for GO:0006006</a> at QuickGO
Subset	gosubset_prok
Related	<a href="#">Link</a> to all genes and gene products annotated to glucose metabolic process. <a href="#">Link</a> to all direct and indirect annotations to glucose metabolic process. <a href="#">Link</a> to all direct and indirect annotations download (limited to first 10,000) for glucose metabolic process.

View this term in QuickGO.



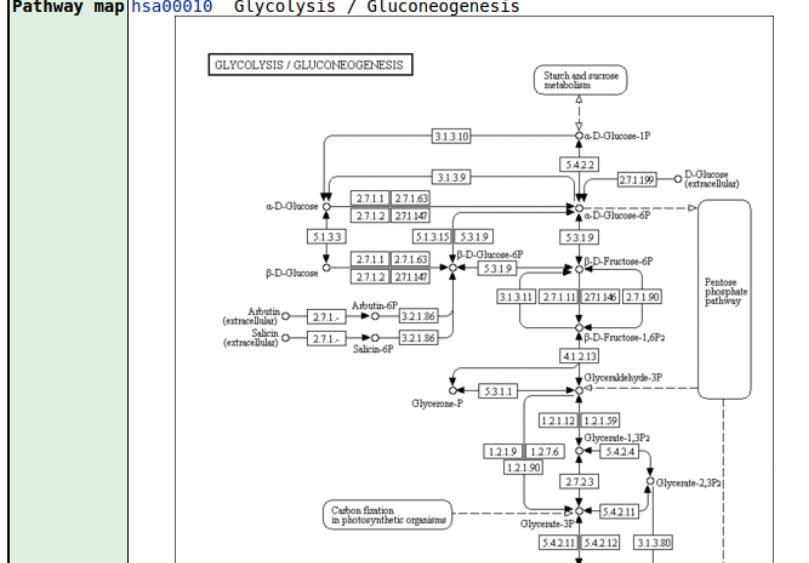
Gene/product	Gene/product name	Annotation qualifier	GO class (direct)	Annotation extension	Contributor	Organism	Evidence
LOC101076683	Uncharacterized protein		glucose metabolic process		GO_Central	Takifugu rubripes	IBA
Soly03g118650.2			glucose metabolic process		GO_Central	Solanum lycopersicum	IBA
Os12g0190000	Os12g0190000 protein		glucose metabolic process		GO_Central	Oryza sativa Japonica Group	IBA
LOC101072156	Uncharacterized protein		gluconeogenesis		GO_Central	Takifugu rubripes	IBA
SDHAF3			regulation of gluconeogenesis		GO_Central	Macaca mulatta	IBA
g6pc3.1	Uncharacterized protein		gluconeogenesis		GO_Central	Xenopus tropicalis	IBA
CHLREDRAFT_113717	Predicted protein		glucose metabolic process		GO_Central	Chlamydomonas reinhardtii	IBA
gdppg1	GDP-D-glucose phosphorylase 1		glucose metabolic process		GO_Central	Xenopus tropicalis	IBA
FBP1			gluconeogenesis		GO_Central	Pan troglodytes	IBA
G6PC2	Uncharacterized protein		gluconeogenesis		GO_Central	Macaca mulatta	IBA

# 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 заболеваниям
  - ✓ ссылкам на литературу и др.

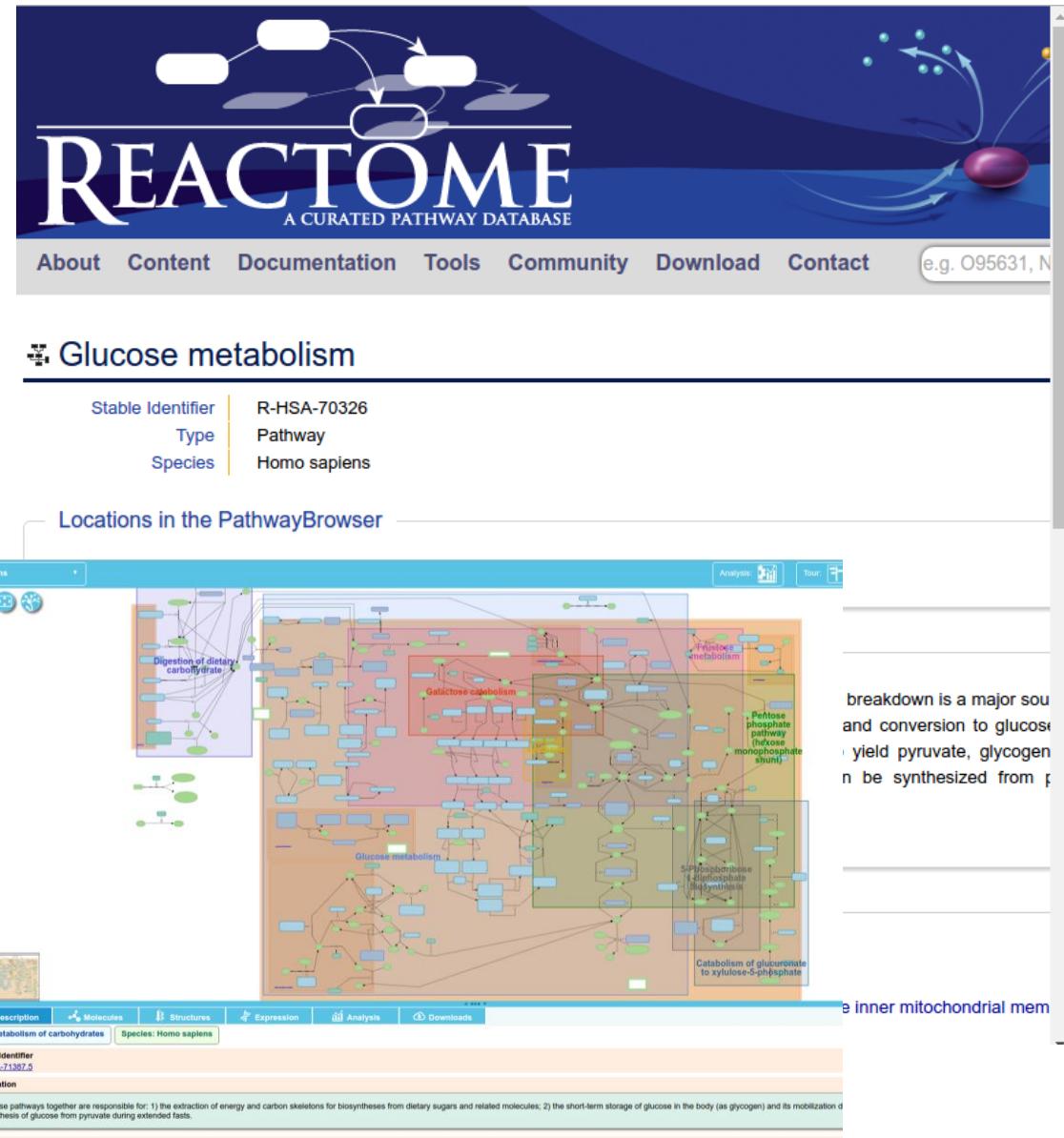
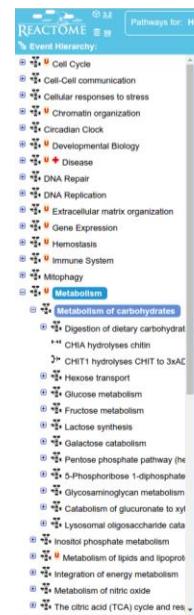
Информационная карточка биологического процесса.

KEGG PATHWAY: hsa00010 Help

Entry	hsa00010	Pathway
Name	Glycolysis / Gluconeogenesis - Homo sapiens (human)	
Description	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, glycerate-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].	
Class	Metabolism; Carbohydrate metabolism BRITE hierarchy	
Pathway map	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.



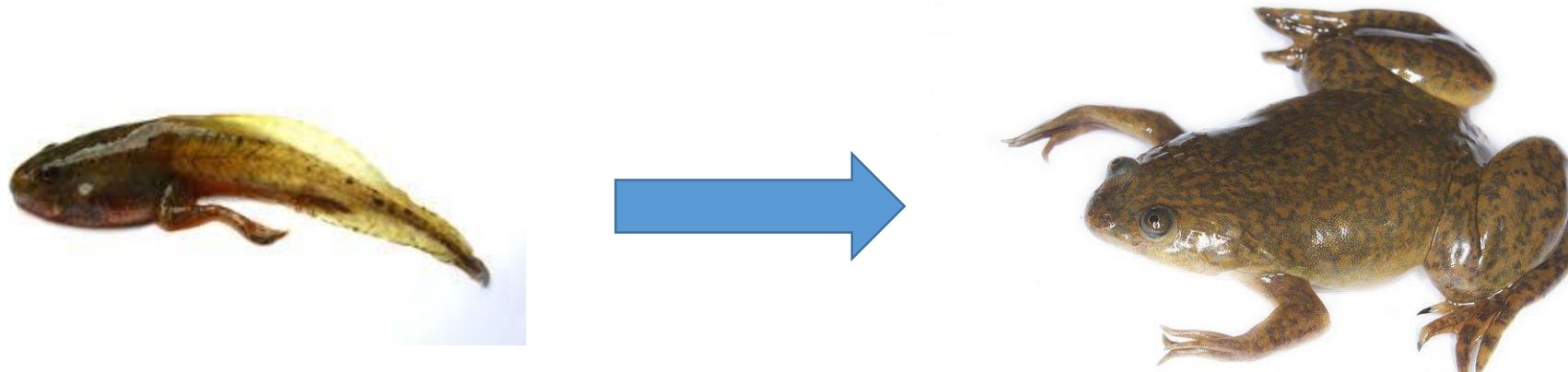
Информационная карточка  
биологического процесса.

Какой процесс отвечает за «потерю хвостика» головастиком?



## Какой процесс отвечает за «потерю хвостика» головастиком?

Апоптоз (др.-греч. ἀπόπτωσις – 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 УУ1

Search NCBI databases

Help

уу1

Search

Results found in 33 databases for "уу1"

## Literature

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

## Health

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

## Genomes

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

## Genes

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

## Proteins

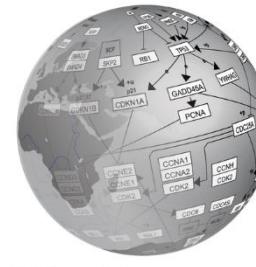
Conserved Domains	7	conserved protein domains
Protein	979	protein sequences
Protein Clusters	1	sequence similarity-based protein clusters
Structure	19	experimentally-determined biomolecular structures

## Chemicals

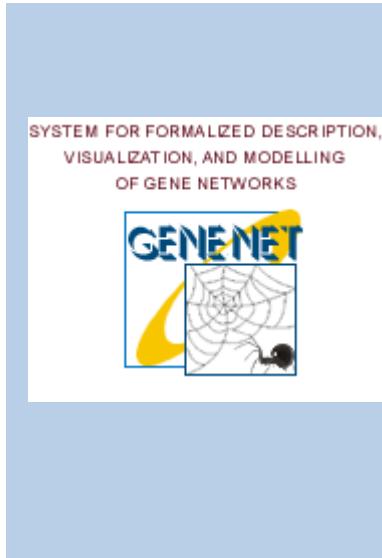
BioSystems	767	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	7	bioactivity screening studies
PubChem Compound	1	chemical information with structures, information and links
PubChem Substance	280	deposited substance and chemical information

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

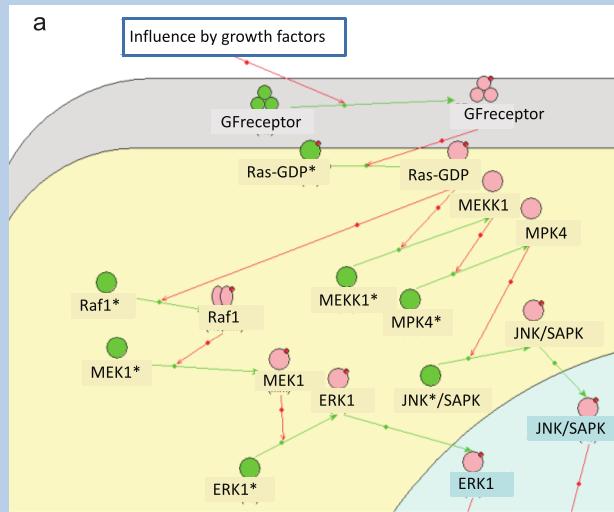
1. Ручная реконструкция генных сетей экспертами с использованием специальных программных средств – редакторов генных сетей.



WIKIPATHWAYS  
Pathways for the People



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, активируемый 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.

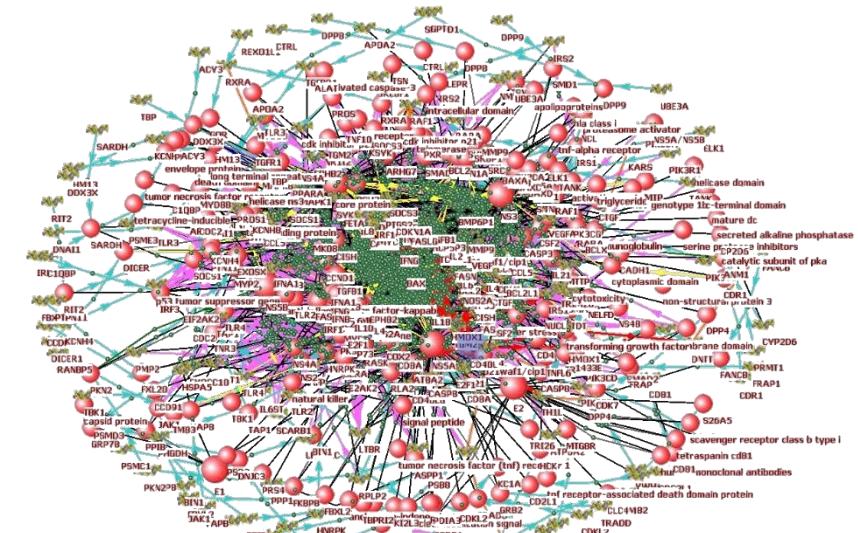
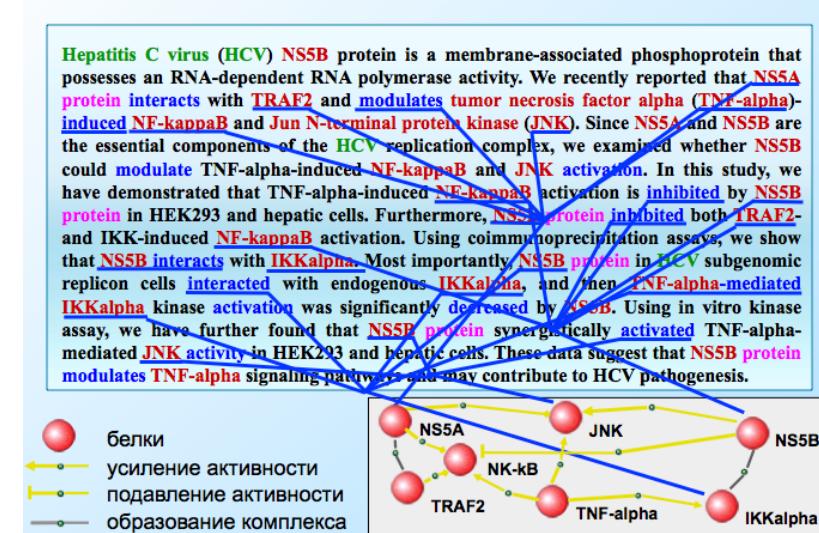
Система ANDSystem включает

- 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.



# 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.

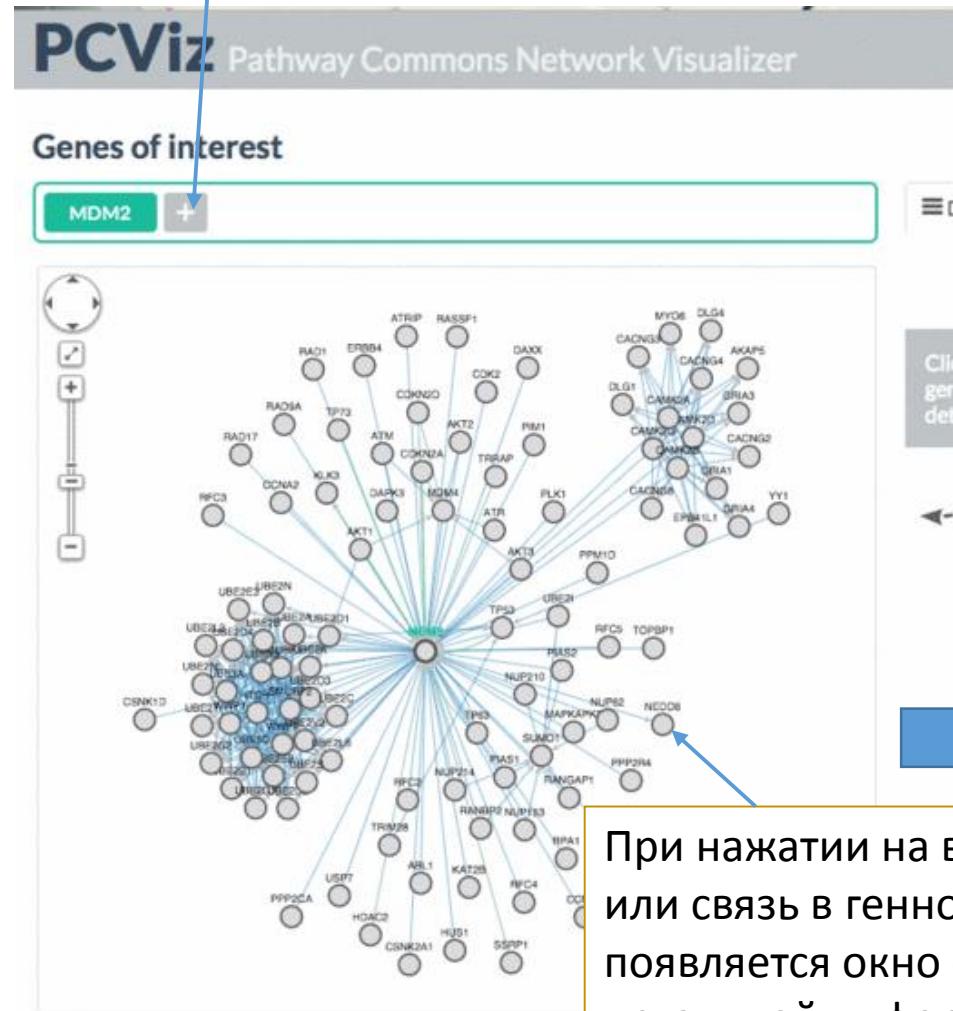
## Pathway Commons интегрирует информацию из 24 баз данных

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

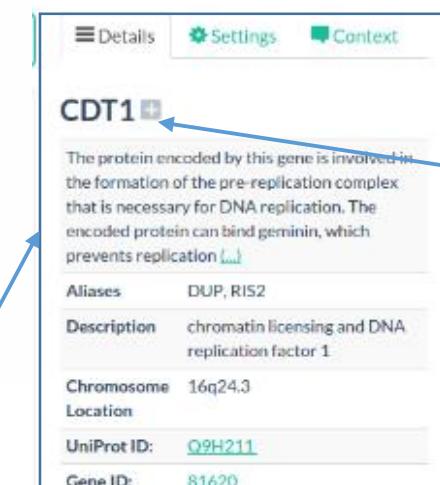
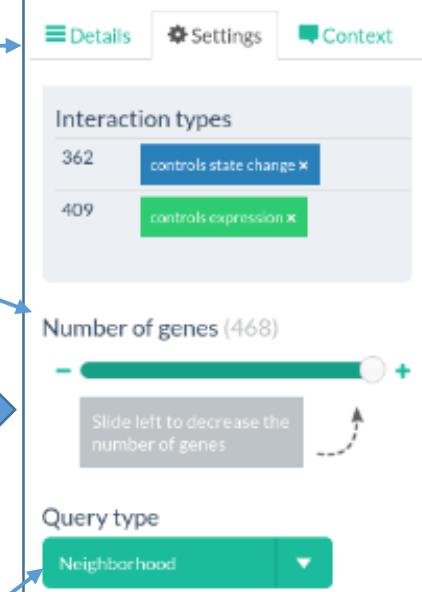
Текстовое поле для ввода  
названия интересующего гена

Добавление в сеть данных по изменениям,  
наблюдающимся при раке

Настройки для фильтрации сети по связям и вершинам



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

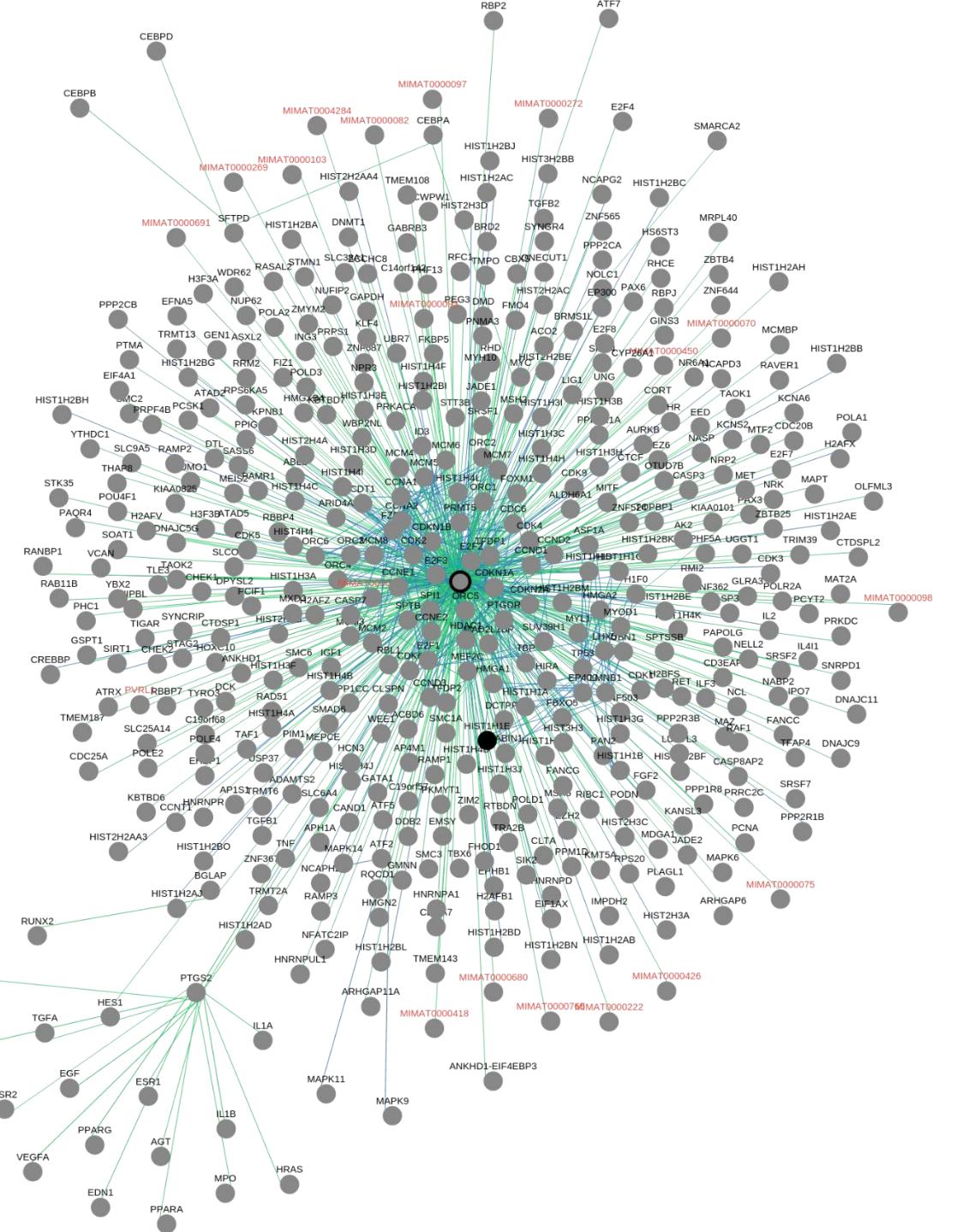


Добавление  
гена в список  
интересующих  
вас генов

# 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

Системная биология и моделирование  
ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Вестникский журнал генетики и геномики. 2016;20(6):840-847  
DOI:10.18699/VJ16200

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

О.В. Саук<sup>1,2</sup>, Н.А. Коновалова<sup>2</sup>, П.С. Деменков<sup>1</sup>, Н.В. Иванченко<sup>2,3</sup>, Т.В. Иванченко<sup>1</sup>, А.В. Иванченко<sup>1</sup>,

М.Н. Пономарева<sup>2</sup>, О.С. Коновалова<sup>2</sup>, О.А. Подольчакина<sup>2</sup>, И.Н. Лаврик<sup>1,4</sup>, Н.А. Колчак<sup>1</sup>, В.А. Иванченко<sup>1</sup>

<sup>1</sup>Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр по этиологии и генетике Себерского отделения Российской академии наук», Новосибирск, Россия

<sup>2</sup>Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный медицинский университет», Министерство здравоохранения Российской Федерации, Томск, Россия

<sup>3</sup>Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный

медицинский университет», Новосибирск, Россия

Saik et al. BMC Medical Genomics 2018, 11(Suppl 1):5  
DOI:10.1186/s12920-018-0331-4

BMC Medical Genomics

## RESEARCH

## Open Access



Novel candidate genes important for  
asthma and hypertension comorbidity  
revealed from associative gene networks

Olga V. Saik<sup>1\*</sup>, Pavel S. Demenkov<sup>1</sup>, Timofey V. Ivanisenko<sup>1</sup>, Elena Yu Bragina<sup>2</sup>, Maxim B. Freidin<sup>2</sup>,

Irina A. Goncharova<sup>2</sup>, Victor E. Dosenko<sup>3</sup>, Olga I. Zolotareva<sup>4</sup>, Ralf Hofstaedt<sup>5</sup>, Inna N. Lavrik<sup>6</sup>,

Eugeny I. Rogov<sup>1,7,8,9</sup> and Vladimir A. Ivanisenko<sup>1</sup>

From Belyaev Conference  
Novosibirsk, Russia. 07-10 August 2017

## Abstract

**Background:** Hypertension and bronchial asthma are a major issue for people's health. As of 2014, approximately one billion adults, or ~22% of the world population, have had hypertension. As of 2011, 235–330 million people globally have been affected by asthma and approximately 250,000–345,000 people have died each year from the disease. The development of the effective treatment therapies against these diseases is complicated by their comorbidity features. This is often a major problem in diagnosis and their treatment. Hence, in this study the bioinformatic methodology for the analysis of the comorbidity of these two diseases have been developed. As such, the search for candidate genes related to the comorbid conditions of asthma and hypertension can help in elucidating the molecular mechanisms underlying the comorbid condition of these two diseases, and can also be useful for genotyping and identifying new drug targets.

**Results:** Using ANDSystem, the reconstruction and analysis of gene networks associated with asthma and hypertension was carried out. The gene network of asthma included 713 genes/proteins and 62,603 interactions, while the gene network of hypertension – 713 genes/proteins and 45,479 interactions. Two hundred and five genes/proteins and 9638 interactions were shared between asthma and hypertension. An approach for ranking genes implicated in the comorbid condition of two diseases was proposed. The approach is based on nine criteria for ranking genes by their importance, including standard methods of gene prioritization (Endeavor, ToppGene) as well as original criteria that take into account the characteristics of an associative gene network and the presence of known polymorphisms in the analysed genes. According to the proposed approach, the genes IL10, TLB4, and CAT had the highest priority in the development of comorbidity of these two diseases. Additionally, it was revealed that the list of top genes is enriched with apoptotic genes and genes involved in biological processes related to the functioning of central nervous system.

(Continued on next page)

## STRING

# SCIENTIFIC REPORTS

## OPEN

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

Juthatip Manissom, Supaporn Khamchun, Arada Vinaphat & Visith Thongboonkerd

Received: 24 February 2016  
Accepted: 10 June 2016  
July 2016

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 of protein-protein interactions. Overexpression of  $\alpha$ -tubulin (pcDNA6.2-TUBA1A) was then performed and its efficacy was confirmed. 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  $\alpha$ -enolase) 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 mechanism remains 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 nuclei" and adhere to already existing crystals on renal tubular epithelial cells<sup>3,4</sup>. Adhesion of crystals onto the cells is a critical event which triggers many cascades of cellular response, 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>7,8</sup>.

Even with the aforementioned knowledge, molecular mechanisms of the downstream cellular response remain largely unknown. From our previous expression proteomics study<sup>9</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, including cell cycle, cell division, cell migration, protein degradation, heat shock, 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 VT.(email: thongboonkerd@dr.com)

## RESEARCH

## Open Access



PLOS COMPUTATIONAL BIOLOGY

CrossMark

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 week 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<sub>2.5</sub> (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<sub>2.5</sub>, 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 interconnected 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 vitro 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 source, provide a link to the Creative Commons license, and indicate if changes were made. No reuse allowed without permission.  
<http://creativecommons.org/publicdomain/zero/1.0/> applies to the data made available in this article, unless otherwise stated.

© The Author(s). 2016 Mapping 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.

## Pathway Commons

### RESEARCH ARTICLE

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

Yasin Şenbaşoğlu<sup>1\*</sup>, Selçuk Onur Sömer<sup>1</sup>, Francisco Sánchez-Vega<sup>1</sup>, Debra Bemis<sup>1</sup>, Giovanni Grilelio<sup>2</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  
[senbay@mksc.org](mailto:senbay@mksc.org) (YS); [csander@mskcc.org](mailto:csander@mskcc.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.

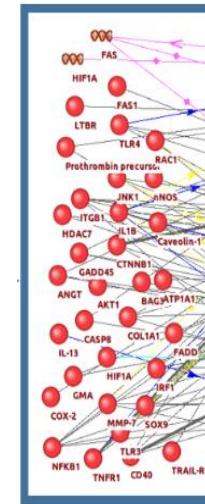
## Author Summary

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

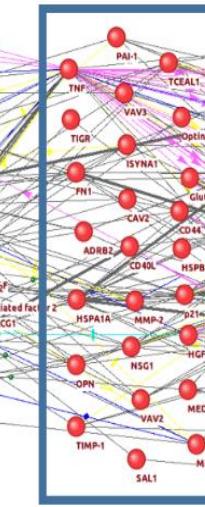
# Реконструкция молекулярно-генетических механизмов взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при первичной открыто-угольной глаукоме с помощью системы ANDSystem.

- 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 нарушенной 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.

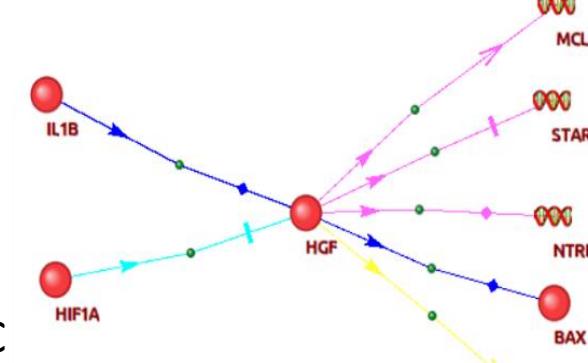
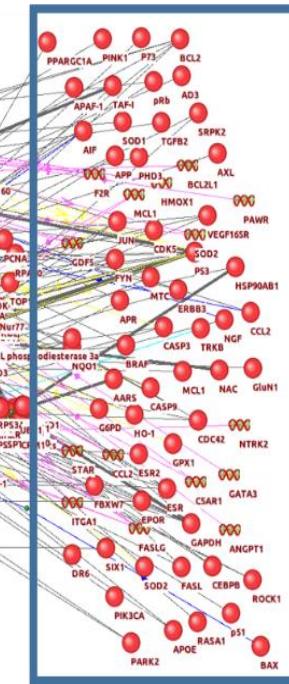
#### Белки/гены, вовлеченные в ОКМС



25 генов/белков-посредников ассоциированных с ПОУГ

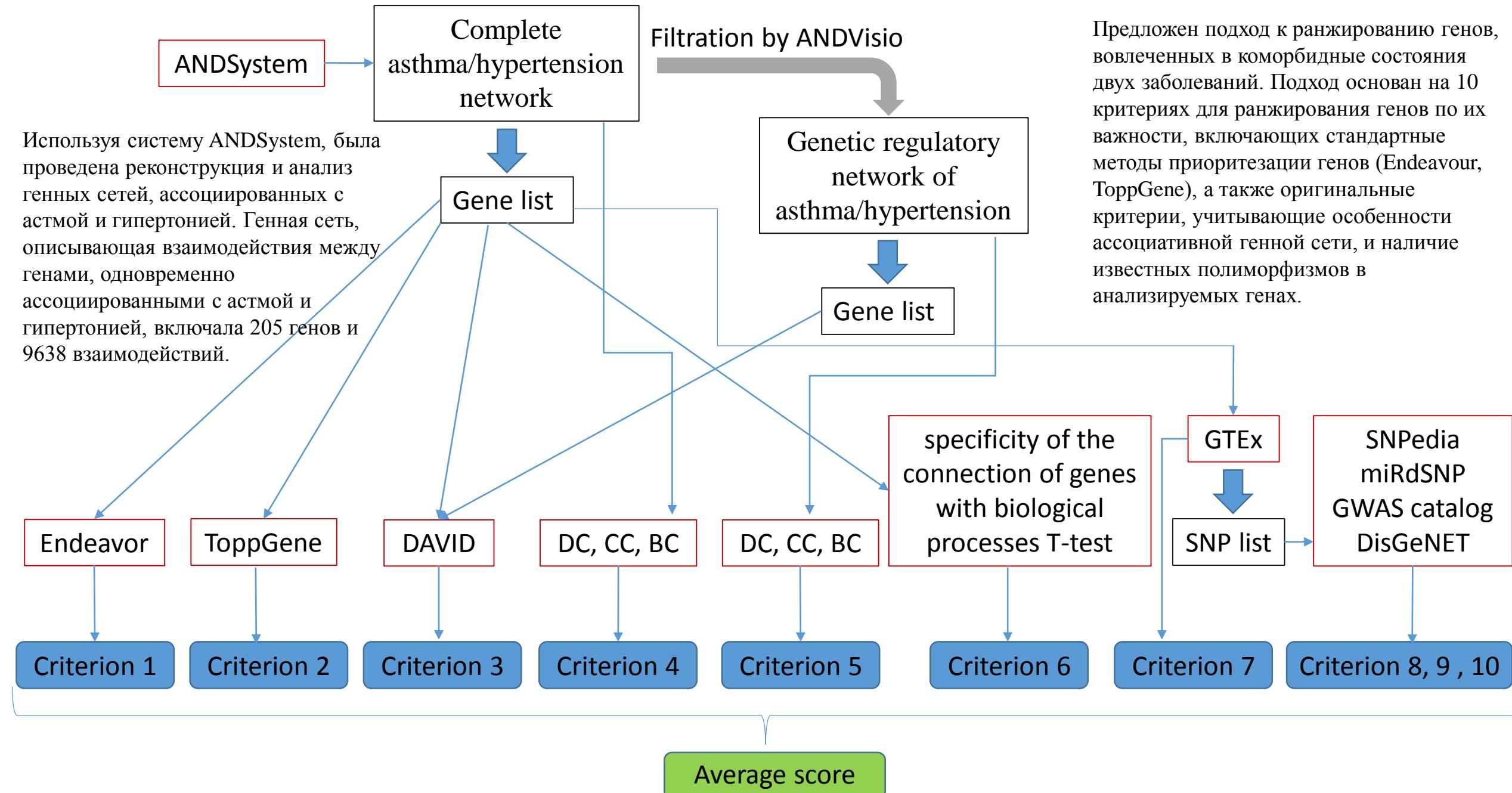


Белки/гены, вовлеченные в нейрональный апоптоз

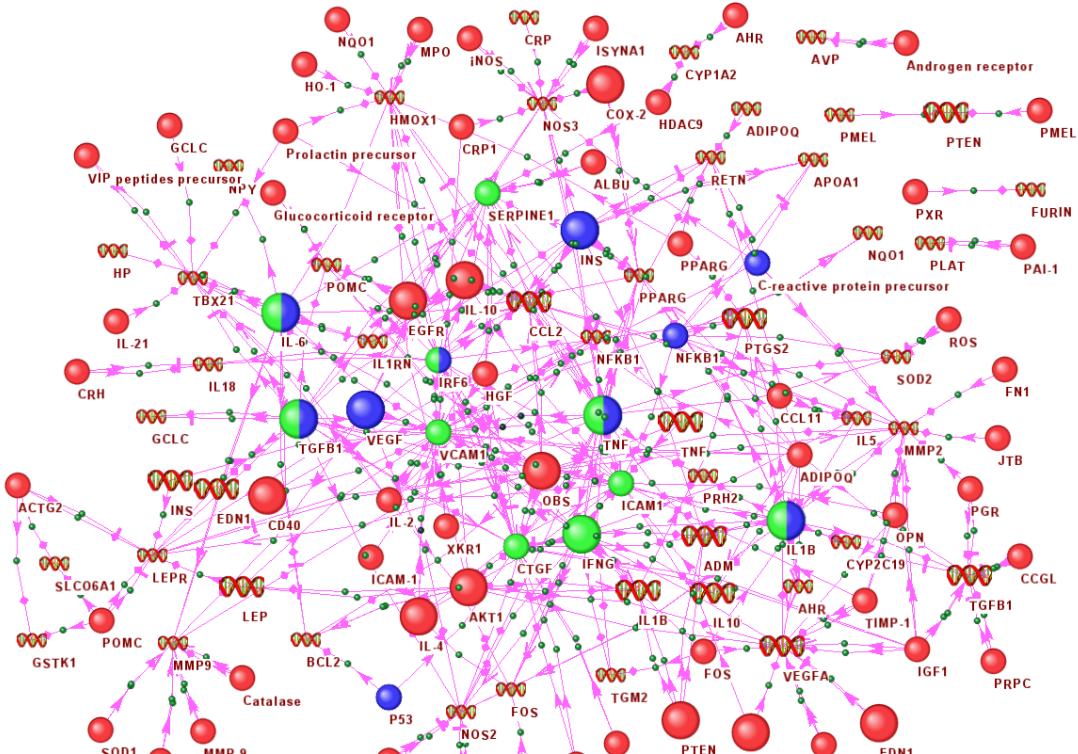


Ассоциативная сеть, описывающая потенциальные взаимодействия между ОКМС и нейрональным апоптозом через молекулярно-генетические цепочки. а) Общая сеть, включающая 25 генов/белков посредников, ассоциированных с ПОУГ. б) Пример цепочек, проходящих через фактор роста гепатоцитов (HGF), выступающий в качестве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

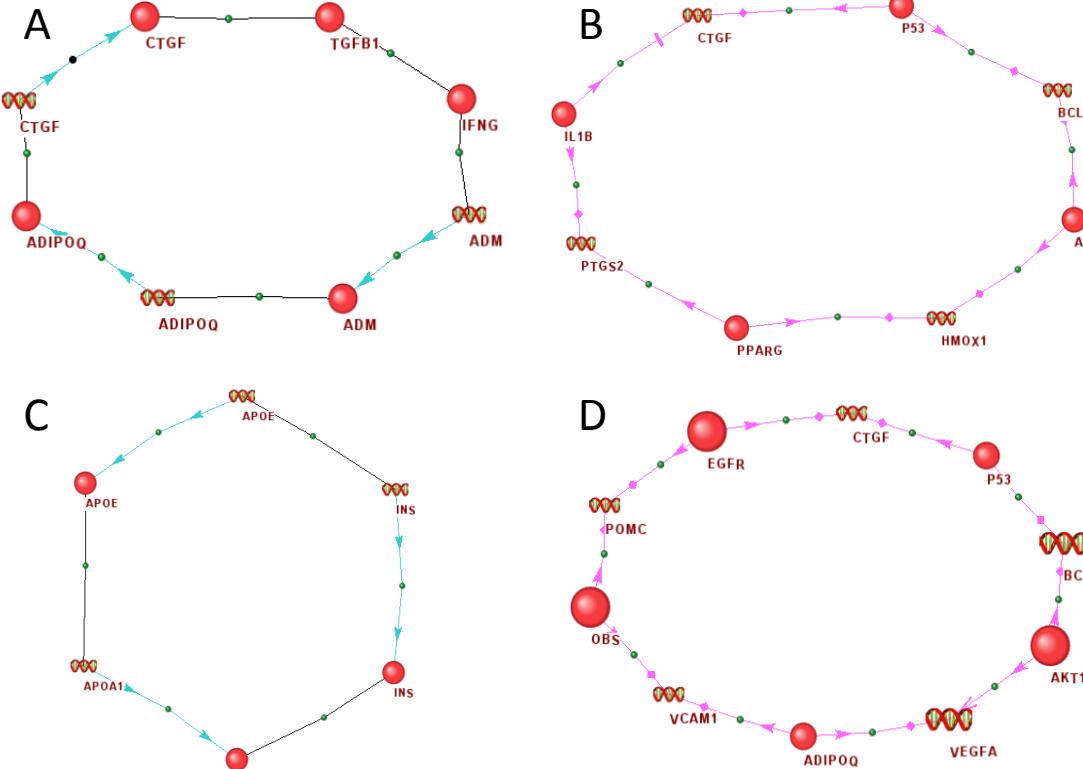
# Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks



Генетически регуляторная сеть астма/гипертония. Крупными шарами показаны белки, вовлеченные в большое число сверхпредставленных Gene Ontology биологических процессов. Топ 10 наиболее центральных белков в полной сети астма/гипертония, выделены синим цветом; в генетически регуляторной сети астма/гипертония -- зеленым цветом.



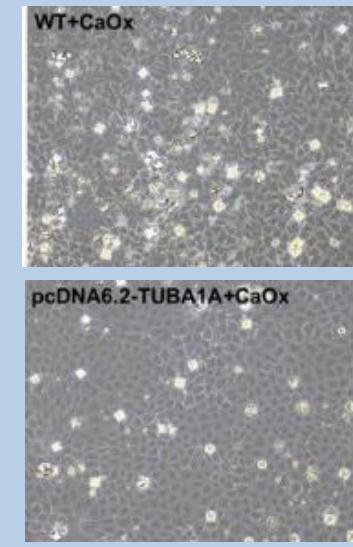
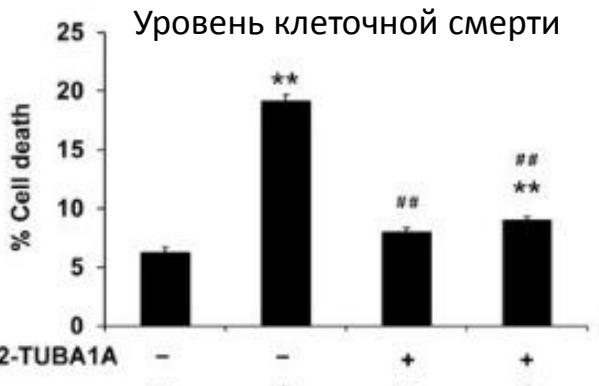
Фундаментальные циклы, выявленные в полной и генетически регуляторной сетях астма/гипертония, связанные с апоптозом и ЦНС. (А) цикл из полной сети астма/гипертония, включающий только гены/белки, ассоциированные с апоптозом. (В) цикл из генетически регуляторной сети астма/гипертония, включающий только гены/белки, ассоциированные с апоптозом. (С) цикл из полной сети астма/гипертония, включающий только гены/белки, ассоциированные с ЦНС. (Д) цикл из генетически регуляторной сети астма/гипертония, включающий 5 участников, ассоциированных с ЦНС.



Согласно предложенному подходу гены IL10, TLR4 и CAT имели наивысший приоритет для сочетанного развития астмы и гипертонии

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

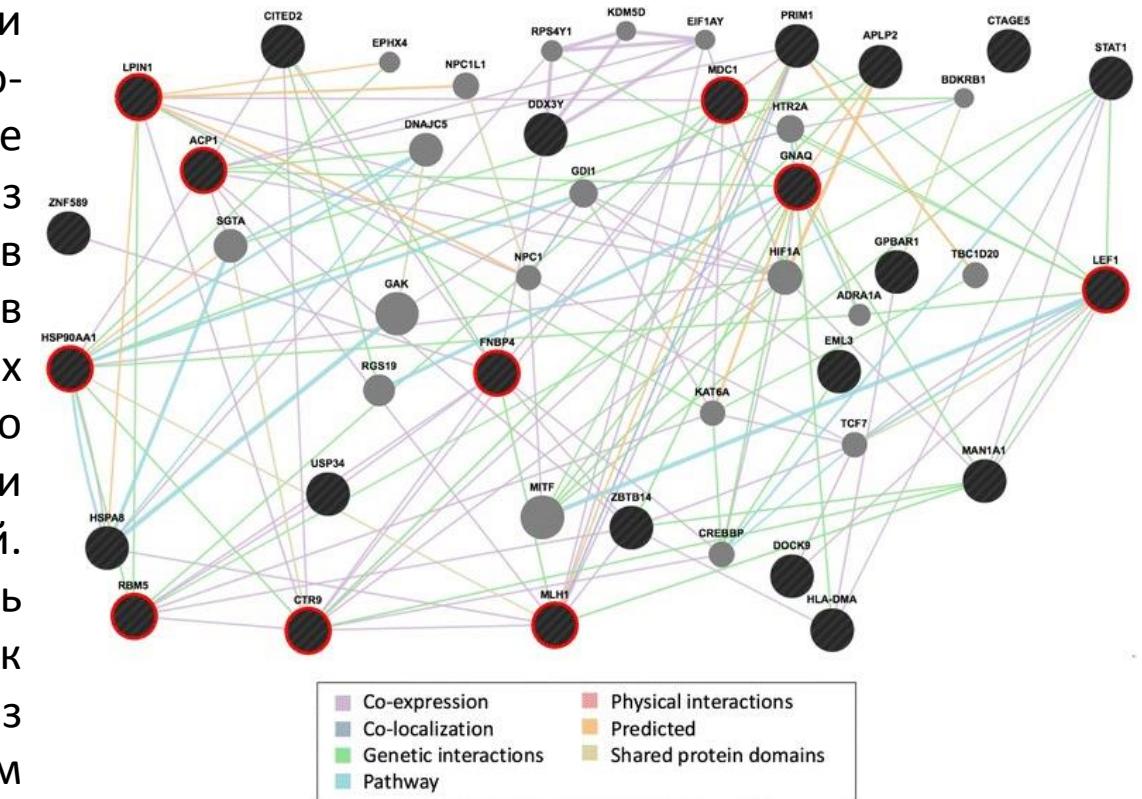
Адгезия кристаллов оксалата кальция (CaOx) на почечных канальцевых эпителиальных клетках наблюдается при мочекаменной болезни. Было показано, что экспрессия ряда белков достоверно меняется в клетках, индуцированных кристаллами CaOx. Анализ глобальной сети белок-белок взаимодействий с использованием системы STRING показал, что а-тубулин, экспрессия которого была снижена в эксперименте, является одним из центральных узлов сети белок-белковых взаимодействий. Для подтверждения ключевой роли а-тубулина была исследована клеточная линия pcDNA6.2-TUBA1A, в которой а-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 а-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.

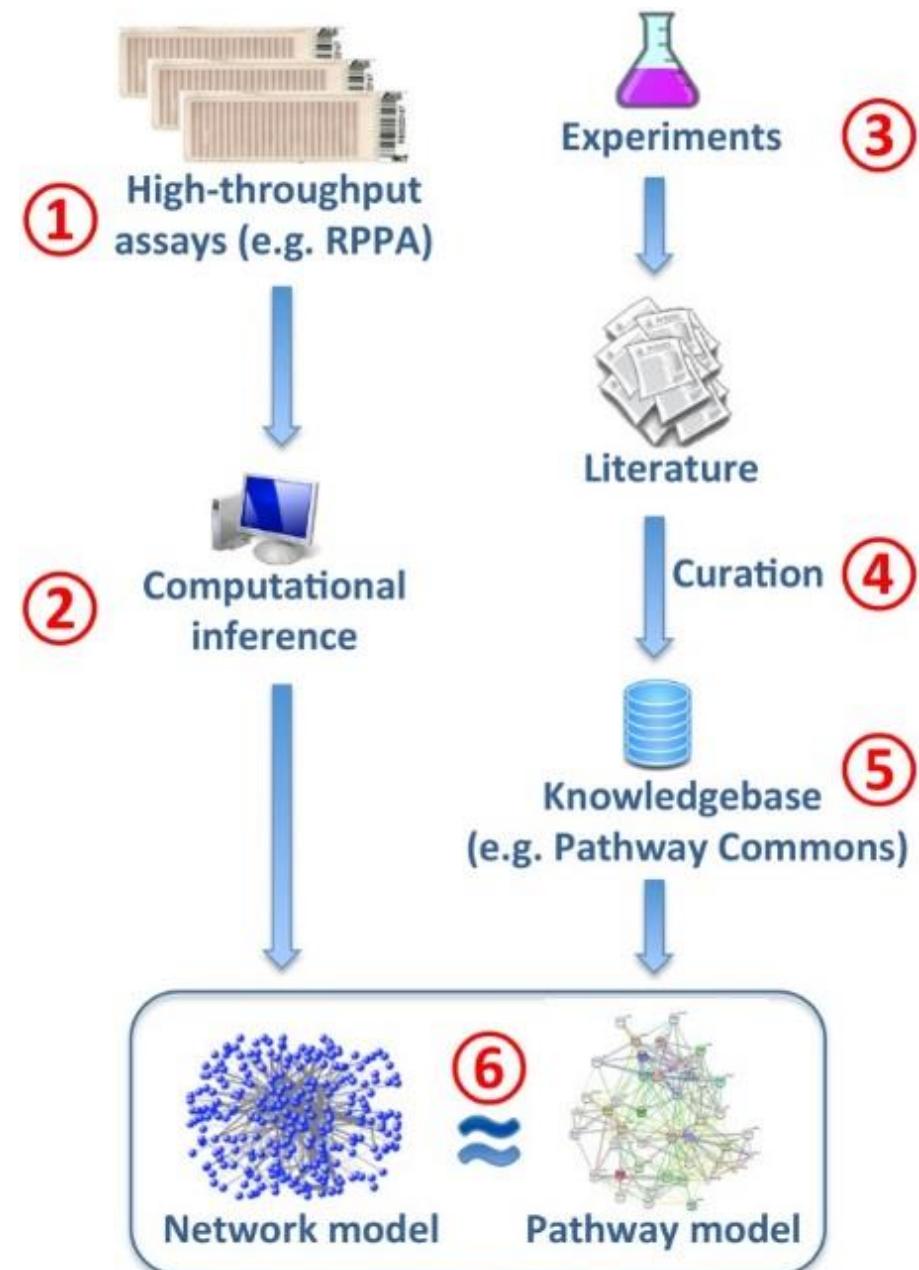
Воздействие загрязненного воздуха связано с неблагоприятными последствиями для здоровья, включая увеличение сердечно-легочной заболеваемости и смертности, а также увеличение риска рака легких. Был проведен полно-геномный анализ уровней экспрессии РНК крови 63 некурящих сотрудников парковочных терминалов 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 взаимодействий 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!*