

Лекция №5

Интеграция разрозненных источников информации по биомедицинской тематике

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

План лекции

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

1. Введение

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

Компоненты генной сети:

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

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

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

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

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

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

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

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

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

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:

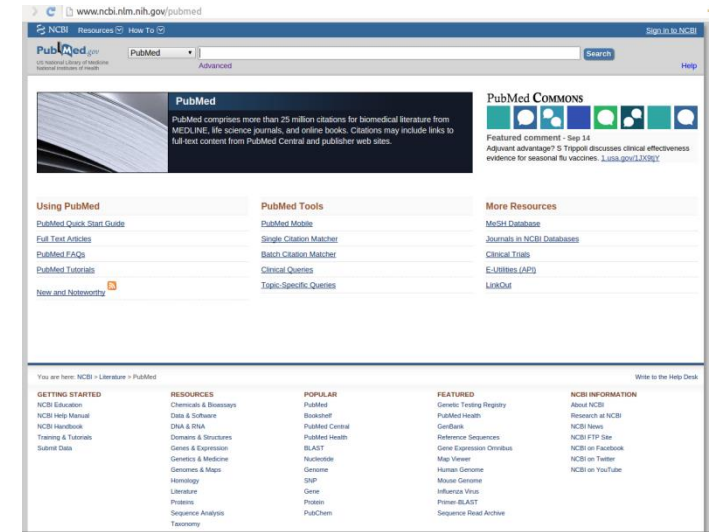
- (1) **молекулярно-генетические объекты**. К этому классу относятся такие типы объектов, как гены, РНК, белки, метаболиты, клеточные компоненты;
- (2) **биологические процессы и системы** (метаболические пути, пути передачи сигналов, транспортные пути и т.д.);
- (3) **фенотипические признаки**, поведенческие 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.gov US National Library of Medicine National Institutes of Health

PubMed *Beta vulgaris* Search

Create RSS Create alert Advanced Help

Article types Clinical Trial Review Customize ...

Text availability Abstract Free full text Full text

PubMed Commons Reader comments Trending articles

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

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](#)

Results by year

Download CSV

Related searches

beta vulgaris antioxidant

NCBI Resources How To Sign In to NCBI

PubMed.gov US National Library of Medicine National Institutes of Health

PubMed *Panax ginseng* Search

Create RSS Create alert Advanced Help

Article types Clinical Trial Review Customize ...

Text availability Abstract Free full text Full text

PubMed Commons Reader comments Trending articles

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

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](#)

Results by year

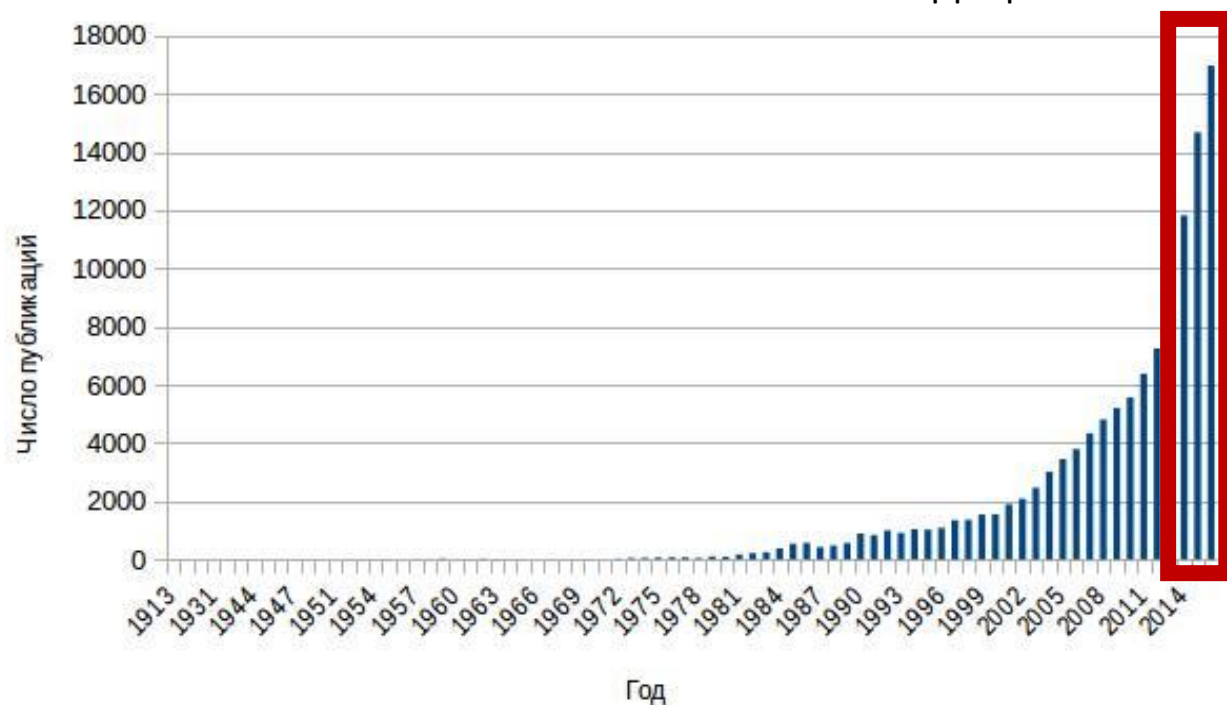
Download CSV

Related searches

panax ginseng review

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

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



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



Китайская традиционная медицина – альтернатива западным методам

11.12.2016 32 1900

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

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

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

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



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

Компоненты генной сети:

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

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

5) И другие биологические объекты

Базы данных, содержащие информацию по генам.



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

HGNC - HUGO Gene Nomenclature Committee

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

The screenshot displays the HGNC website interface. At the top, the URL is www.genenames.org/cgi-bin/gene_symbol_report?hgnc_id=HGNC:12856. The HGNC logo and name are visible. A search bar contains the text "Search everything" and "Search symbols, keywords or IDs". Below the search bar is a navigation menu with links: Home, Downloads, Gene Families, Tools, Useful Links, About, Newsletters, Contact Us, Help, VGNC, and Request Symbol. The main content area is titled "Symbol Report: YY1". It lists various attributes for the gene YY1, including its approved symbol, name, HGNC ID, previous symbols, locus type, chromosomal location, gene family, and HCOPI. A table of homologs is also shown, listing Mus musculus and Rattus norvegicus. Below the homologs table, there are sections for Gene Resources, Nucleotide Sequences, Protein Resources, and Clinical Resources, each with links to external databases and resources.

Symbol Report: YY1

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
HCOPI ? Orthology Predictions for YY1

External links

HOMOLOGS ?

	Symbol	Database
Mus musculus	Yy1	MGI:99150 C
Rattus norvegicus	Yy1	RGD:3982 D

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

NUCLEOTIDE SEQUENCES ? BC020324 C [GenBank](#) [ENA](#) [DDBJ](#)
NM_003403 C [RefSeq](#) [NCBI Sequence Viewer](#)
CCDS9957 C [CCDS](#)

PROTEIN RESOURCES ? P25490 D [UniProt](#) [InterPro](#) [PDBe](#)

CLINICAL RESOURCES ? OMIM: [600013](#) D
[GeneTests](#) D
[Orphanet](#) D
DECIPHER D

NCBI Gene

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

Поиск

Окно для ввода запросов

запросов

Сохранение

Организмы

The screenshot shows the NCBI Gene database search results for the query 'yy1'. The search bar at the top contains 'yy1' and a 'Search' button. Below the search bar, there are options for 'Gene sources', 'Categories', 'Sequence content', and 'Status'. The search results are displayed in a table with columns for Name/Gene ID, Description, Location, and Aliases. Four specific results are highlighted with red boxes and blue arrows pointing from the text on the left: YY1 (ID: 7528) in Homo sapiens, YY1 (ID: 22632) in Mus musculus, YY1 (ID: 24919) in Rattus norvegicus, and YY1 (ID: 826093) in Arabidopsis thaliana. The right sidebar shows filters for 'Message Filters' and 'Results by taxon', with a list of top organisms including Homo sapiens (235), Mus musculus (66), Rattus norvegicus (25), Cyprinus carpio (13), and Sinocyclocheilus anshuiensis (10). There is also a 'Search details' section with a search query 'yy1[All Fields] AND alive[prop]' and a 'Recent activity' section showing 'yy1 AND (alive[prop]) (1415)' and 'CAT2 AND (alive[prop]) (1986)'.

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

NCBI Gene. Информационная карточка гена.

Идентификатор

Gene

Full Report

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

Table of contents

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

Genome Browsers

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

Общая информация

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

Более детальная информация

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

NCBI Nucleotide

- Доступна через интернет по адресу <https://www.ncbi.nlm.nih.gov/nucleotide>
- Коллекция нуклеотидных последовательностей из нескольких источников GenBank, RefSeq, TPA и PDB.
- Информация может быть скачана через ftp сайт.
- Идентификаторы обычно начинаются с двух заглавных букв, далее нижнее подчеркивание и число. Например, одна из последовательностей гена транскрипционного фактора YY1 человека имеет идентификатор NM_003403.4, другая - NG_046908.1, мыши – NM_009537.3

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

Item	Accession	GI	Length	Source
1. Synthetic construct Homo sapiens clone ccsbBroadEn_07141 YY1 gene, encodes complete protein	KJ897747.1	649121807	1,374 bp linear other-genetic	GenBank
2. Homo sapiens YY1 transcription factor (YY1), mRNA	NM_003403.4	459683878	3,159 bp linear mRNA	GenBank
3. Rattus norvegicus YY1 transcription factor (Yy1), mRNA	NM_173290.1	27545349	1,236 bp linear mRNA	GenBank
4. Mus musculus YY1 transcription factor (Yy1), mRNA	NM_009537.3	118130338	2,324 bp linear mRNA	GenBank
5. Danio rerio YY1 transcription factor a (yy1a), mRNA	NM_212617.1	47086800	2,903 bp linear mRNA	GenBank
6. Xenopus laevis YY1 transcription factor L homeolog (yy1.L), mRNA	NM_001087615.1	148234201	1,924 bp linear mRNA	GenBank
7. PREDICTED: Canis lupus familiaris YY1 transcription factor (YY1), mRNA			2,220 bp linear mRNA	

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

NCBI Nucleotide. Карточка последовательности гена.

Особенности первичной структуры последовательности

NCBI Reference Sequence: NM_003403.4

Homo sapiens YY1 transcription factor (YY1), mRNA

NCBI Reference Sequence: NM_003403.4

FASTA Graphics

Go to: []

LOCUS NM_003403 3159 bp mRNA linear PRI 07-OCT-2016

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

ACCESSION NM_003403

VERSION NM_003403.4

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

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

REFERENCE 1 (bases 1 to 3159)

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

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

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

PUBMED 26800240

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

REFERENCE 2 (bases 1 to 3159)

AUTHORS Nieborak A and Gorecki A.

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

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

PUBMED 26910132

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

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the YY1 gene

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

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

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

See all...

Pathways for the YY1 gene

UCH proteinases

Deubiquitination

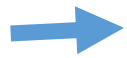
TFAP2 (AP-2) family regulates transcription of growth factors and their receptors

See all...

```
exon 1384..1542
      /gene="YY1"
      /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
      /inference="alignment:Splign:2.0.8"
exon 1543..3159
      /gene="YY1"
      /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
      /inference="alignment:Splign:2.0.8"
STS 1729..1852
     /gene="YY1"
     /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
     /standard_name="RH1618"
     /db_xref="UniSTS:42115"
STS 1954..2640
     /gene="YY1"
     /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
     /standard_name="YY1_3892"
     /db_xref="UniSTS:462950"
STS 2035..2173
     /gene="YY1"
     /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
     /standard_name="G06170"
     /db_xref="UniSTS:49520"
regulatory 2690..2695
           /regulatory_class="polyA_signal_sequence"
           /gene="YY1"
           /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
polyA_site 2714
           /gene="YY1"
           /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
STS 2879..3086
     /gene="YY1"
     /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
     /standard_name="RH68412"
     /db_xref="UniSTS:51218"
STS 3001..3100
     /gene="YY1"
     /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
     /standard_name="RH44921"
     /db_xref="UniSTS:36888"
regulatory 3106..3111
           /regulatory_class="polyA_signal_sequence"
           /gene="YY1"
           /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
polyA_site 3132
           /gene="YY1"
           /gene_synonym="DELTA; IN080S; NF-E1; UCRBP; YIN-YANG-1"
ORIGIN
1 agggcgaacg ggcgagtggc agcgaggcgg ggcgggctga ggccagcggc gaagtctcgc
61 gaggccggcg ccgagcagag tgtggcggcg gcggcgagat ctgggctcgg gttgaggagt
121 tggattttgt gtggaaggag gcggaaggcg aggaggaagg ggaagcggga ggcgccggcc
181 ggaggcggcg aggagggcgg gccaggcggg gcggttgcgg cgaagcgagg cgaggcgggg
241 agccgagacg agcagcggcg gacgagcggc gcaccgaggc gacggaggcg
301 gggaagcccc gccgcccggc cggcgcccgc cccttcccc gccgcccgcc ccctctcccc
361 ccgcccgctc gccgcttcc tcctctgcgc ttcttcccc acggccggcc gccctctcgc
421 ccgccggccc gcagccgagg agccgagggc gccgcgggcg tggcggcggg gccctcagcc
```

Общая информация

Локус
Идентификатор
Организм и
таксономия
Ссылки на
литературу



Продолжение информационной карточки

GeneCards

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

Информационная карточка гена.

The screenshot displays the GeneCards website interface for the YY1 gene. The page title is "YY1 Gene (Protein Coding) ★" with the subtitle "YY1 Transcription Factor". The GCID is GC14P100238 and there are 64 GIFs. The page includes a navigation menu with options like Home, User Guide, Analysis Tools, News And Views, About, My Genes, and Log In / Sign Up. Below the title, there are tabs for various gene-related topics: Aliases, Disorders, Domains, Drugs, Expression, Function, Genomics, Localization, Orthologs, Paralogs, Pathways, Products, Proteins, Publications, Sources, Summaries, Transcripts, and Variants. The page also features advertisements for EMD Millipore, GenScript, ORIGENE, and Vigene. A section titled "Aliases for YY1 Gene" lists various names like YY1 Transcription Factor, INO80 Complex Subunit S, Delta Transcription Factor, Yin And Yang 1 Protein, INO80S, NF-E1, and YY-1. Another section lists "External Ids for YY1 Gene" including HGNC, Entrez Gene, Ensembl, OMIM, and UniProtKB. A "Previous GeneCards Identifiers for YY1 Gene" section lists several GC IDs. The "Summaries for YY1 Gene" section includes an "Entrez Gene Summary for YY1 Gene" which states: "YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins. The protein is involved in repressing and activating a diverse number of promoters. YY1 may direct histone deacetylases and histone acetyltransferases to a promoter in order to activate or repress the promoter, thus implicating histone modification in the function of YY1. [provided by RefSeq, Jul 2008]"

Ensembl

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

Информационная карточка гена.

The screenshot displays the Ensembl gene page for YY1 (ENSG00000100811). The page is organized into several sections:

- Gene-based displays:** A sidebar menu with categories like Summary, Sequence, Comparative Genomics, Ontologies, Genetic Variation, and ID History.
- Gene: YY1 ENSG00000100811:** The main header for the gene.
- Description:** YY1 transcription factor [Source:HGNC Symbol;Acc:HGNC:12856].
- Synonyms:** INO80S, hsa-mir-6764, UCRBP, YIN-YANG-1, NF-E1, DELTA.
- Location:** Chromosome 14: 100,238,298-100,282,792 forward strand. GRCh38:CM000676.2.
- About this gene:** This gene has 7 transcripts (splice variants), 63 orthologues, 17 paralogues, is a member of 1 Ensembl protein family and is associated with 1 phenotype.
- Transcripts:** A button to "Show transcript table".
- Summary:** A section with a question mark icon.
- Name:** YY1 (HGNC Symbol).
- CCDS:** This gene is a member of the Human CCDS set: CCDS9957.1.
- UniProtKB:** This gene has proteins that correspond to the following UniProtKB identifiers: P25490.
- RefSeq:** Overlapping RefSeq Gene ID 7528 matches and has similar biotype of protein_coding.
- Ensembl version:** ENSG00000100811.11.
- Other assemblies:** This gene maps to 100,704,635-100,749,129 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000100811.
- Gene type:** Known protein coding.
- Annotation method:** Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.
- Alternative genes:** This gene corresponds to the following database identifiers: Havana gene: OTTHUMG00000150479.
- Annotation Attributes:** overlapping locus [Definitions].

At the bottom, there is a "Go to Region In Detail" button and a genomic track visualization showing the gene structure with exons and introns, and a "Drag/Select" tool.

UCSC (University of California, Santa Cruz) Genome Browser

Информационная карточка гена.

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

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

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

Description: Homo sapiens YY1 transcription factor (YY1), mRNA. (from RefSeq NM_003403)

RefSeq Summary (NM_003403): YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins. YY1 may direct histone deacetylases and histone H3 to a promoter in order to activate or repress the promoter, thus implicating histone modification in the function of YY1. [provided by RefSeq, Jul 2008]. Sequence Note: This RefSeq record was created from transcript and genomic sequence data to make the sequence consistent with the reference genome assembly. The genomic coordinates used for the transcript record were based on transcript alignments. Public RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications. ##Evidence-Data-START## Transcript exon combination :: M77698.1, BC037308.1 [ECO:0000332] RNAseq introns :: single sam

GeneCode Transcript: ENST00000262238.8

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

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

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

Data last updated: 2016-03-28

Sequence and Links to Tools and Databases

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

Comments and Description Text from UniProtKB

ID: [YY1_HUMAN](#)

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

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

GEO - Gene Expression Omnibus

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

Информационная карточка эксперимента.

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

Series GSE14964 Query DataSets for GSE14964

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

Platforms (1): GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array

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

Relations: BioProject PRJNA111841

Analyze with GEO2R

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

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

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

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

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

Пример информационной карточки в UCSC genome browser.

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

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

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

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

RefSeq Summary (NM_002430): Meningioma 1 (MN1) contains two sets of CAG repeats. It is disrupted by a balanced translocation (4;22) in a meningioma, and its inactivation may contribute to meningioma 32 pathogenesis. [provided by RefSeq, Jul 2008]. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications. ##Evidence-Data-START## Transcript exon combination :: X82209.2, CX866728.1 [ECO:0000332] RNAseq introns :: single sample supports all introns SAMEA2147975, SAMN03465404 [ECO:0000348] ##Evidence-Data-END##

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

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

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 pathogenesis. [provided by RefSeq, Jul 2008]. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene page to access additional publications. ##Evidence Data START## Transcript over combination: X82200.2, X866728.1, F50002221, ENAseq intron u...

single sampl

Transcript (

Position:

Coding Reg

Position:

Known Genes

MN1 (uc003adj.3) at chr22:27748277-27801498	- Homo sapiens meningioma (disrupted in balanced translocation) 1 (MN1), mRNA. (from RefSeq NM_002430)
MN1 (uc062csm.1) at chr22:27750064-27796896	- The sequence shown here is derived from an Ensembl automatic analysis pipeline and should be considered as preliminary data. (from UniProt H7C105)
MN1 (uc010gvg.6) at chr22:27750678-27791883	- meningioma (disrupted in balanced translocation) 1 (from HGNC MN1)

Page Index

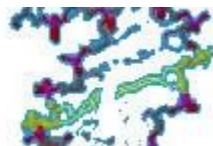
Gene Allele

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

Data last updated: 2013-06-14

Базы данных, содержащие информацию о полиморфизмах в генах.

dbSNP
Short Genetic Variations



 NCBI

SNPedia



GWAS Catalog

- dbSNP
- ClinVar
- SNPedia
- GWAS
- И другие

dbSNP

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

Информационная карточка полиморфизма.

The screenshot displays the dbSNP website interface. At the top, the NCBI logo and 'dbSNP Short Genetic Variations' are visible. A search bar is present with 'dbSNP' selected. The main content area shows the 'Reference SNP (refSNP) Cluster Report: rs61992955'. The report includes a table with the following information:

RefSNP		Allele	
Organism:	human (<i>Homo sapiens</i>)	Variation Class:	SNV: single nucleotide
Molecule Type:	Genomic	RefSNP Alleles:	C/G (FWD)
Created/Updated in build:	129/149	Allele Origin:	
Map to Genome Build:	108/Weight 1	Ancestral Allele:	C
Validation Status:		Variation Viewer:	
		Clinical Significance:	NA
		MAF/MinorAlleleCount:	C=0.0300/150

Below the table, there are links for 'GeneView', 'Map', 'Submission', 'Fasta', 'Resource', 'Diversity', and 'Validation'. The 'Integrated Maps' section shows a table with columns for Assembly, Annotation Release, Chr, Chr Pos, and Contig. Two entries are visible:

Assembly	Annotation Release	Chr	Chr Pos	Contig
GRCh38.p7	108	14	100274488	NT_026437.13
GRCh37.p13	105	14	100740825	NT_026437.12

The 'GeneView' section shows 'GeneView via analysis of contig annotation: YY1 YY1 transcription factor'. It includes a 'Clinical Source' section with radio buttons for 'in gene region', 'cSNP', 'has frequency', and 'double hit'. Below this are two tables: 'Primary Assembly Mapping' and 'RefSeqGene Mapping'.

Assembly	SNP to Chr	Chr	Chr position
GRCh38.p7	Fwd	14	100274488

RefSeqGene	Gene (ID)	SNP to RefS
NG_046908.1	YY1 (7528)	Fwd

The 'Function class:' section is partially visible at the bottom.

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.

Надежный | <https://www.ncbi.nlm.nih.gov/clinvar/variation/91950/#supporting-observation>

NCBI Resources How To Sign in to

ClinVar ClinVar Search ClinVar for gene symbols, HGVS expressions, conditions Advanced

Home About Access Help Submit Statistics FTP

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)

NM_003403.4(YY1):c.1115C>G (p.Thr372Arg)

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

Sample count	Called variants	Potential variants
	no data	0 of 4090

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

← → ↻ <https://www.snpedia.com/index.php/Rs2494262> ☆ ⋮

SNPedia

русский Обсуждение Вклад Создать учётную запись Войти

Искать в SNPedia 🔍

Навигация ▾

Статья Обсуждение Редактировать с помощью формы Править История

Have questions? Visit <https://www.reddit.com/r/SNPedia>

rs2494262

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

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

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

[PMID 25412950] Association of FCER1A genetic polymorphisms

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

? (A;A) (A;C) (C;C) 28

Population	(A;A)	(A;C)	(C;C)
CEU	~35	~45	~20
HCB	~55	~35	~10
JPT	~65	~25	~10
YRI	~10	~15	~75
ASW	~25	~35	~40
CHB	~55	~35	~10
CHD	~55	~35	~10
GIH	~45	~35	~20
LWK	~10	~15	~75
MEX	~10	~15	~75
MKK	~35	~45	~20
TSI	~35	~45	~20
AVG	~25	~35	~40

GWAS Catalog

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

Информационная карточка полиморфизма.

The screenshot shows the GWAS Catalog search results for the SNP rs2766692. The page includes a search bar with the query 'rs2766692' and a search button. Below the search bar, there are navigation links for 'Home', 'Search', 'Diagram', 'Download', 'Documentation', and 'About'. The main content area displays the title 'GWAS Catalog' and the subtitle 'The NHGRI-EBI Catalog of published genome-wide association studies'. A search bar contains the query 'rs2766692' and a search button. Below the search bar, there are examples of search terms: 'breast cancer, rs7329174, Yang, 2q37.1, HBS1L'. The search results are displayed in a table format. The table has columns for 'Author', 'Date', 'Journal', 'Title', 'Reported trait', and 'Association count'. The first result is for Kang SJ (PMID: 22554406) from Genes Brain Behav, published in 2012-05-03. The title is 'Family-based genome-wide association study of frontal θ oscillations identifies potassium channel gene KCNJ6'. The reported trait is 'Electroencephalographic traits in alcoholism' and the association count is 8. There is a '+' button next to the association count. Below the table, there are sections for 'Refine search results' and 'Associations'. The 'Refine search results' section includes options to show results for 'Studies' (1), 'Associations' (1), and 'Catalog traits' (9). It also includes filters for 'p-value' (≤ 5 x 10⁻⁸), 'Odds ratio' (from to), and 'Beta coefficient'. The 'Associations' section includes a table with columns for 'RAF', 'p-value', 'OR', 'Beta', 'CI', 'Region', 'Functional', and 'Reported gene(s)'. The 'Reported gene(s)' column is currently empty.

GWAS / Search / rs2766692

Refine search results

Show results for

- Studies 1
- Associations 1
- Catalog traits 9

Filter results by

p-value ≤ 5 x 10⁻⁸

Odds ratio from to

Beta coefficient

Search results for rs2766692

Download association results

Expand all studies

Studies

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

Associations

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

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

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

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

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

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: AGBL4

Display Settings: Summary, 20 per page, Sorted by SNP_ID

Search results: 1 item: 1 to 20

rs17378497 has merged into rs355206 [Homo sapiens]

1. ATAAAACTGCTTTTCAGTTAAGG[A/C]AAACCTGCTTTTGACCCCTTCCAAA
Chromosome: 1:49206502
Gene: AGBL4 (GeneView)
Functional Consequence: intron variant
Validated: by 1000G, by 2hit 2allele, by cluster, by frequency
Global MAF: A=0.4091/2049
HGVS: NC_000001.10:g.49672174C>A, NC_000001.11:g.49206502C>A, NM_001323573.1:c.413+39268G>T, NM_001323574.1:c.413+39268G>T, NM_001323575.1:c.377+39268G>T, 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

rs17379875 has merged into rs4926831 [Homo sapiens]

2. TGCTAAGCTCTGTTTGAAGCCTCA[C/T]GATTTGTCAAATGATGAGATATGG
Chromosome: 1:49824429
Gene: AGBL4 (GeneView)
Functional Consequence: intron variant
Validated: by 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap
Global MAF: T=0.2883/1444
HGVS: NC_000001.10:g.50290101C>T, NC_000001.11:g.49824429C>T, NM_001323573.1:c.157+26967G>A, NM_001323574.1:c.157+26967G>A, NM_001323575.1:c.157+26967G>A, 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

SNP: GALE

Display Settings: Summary, 20 per page, Sorted by SNP_ID

Search results: 1 item: 1 to 20

rs28940885 [Homo sapiens]

1. CCCAGCTGGCCCAAGGAGGCTGG[A/G]GTGGACAGCAGCCTTAGGGCTGGAC
Chromosome: 1:23796183
Gene: GALE (GeneView)
Functional Consequence: missense
Allele Origin: G(germline)/A(germline)
Clinical significance: Pathogenic
Validated: by 1000G, by cluster, by frequency
Global MAF: T=0.0020/10
HGVS: NC_000001.10:g.24122673C>T, NC_000001.11:g.23796183C>T, NG_007068.1:g.9622G>A, NM_000403.3:c.956G>A, NM_001008216.1:c.956G>A, NM_001127621.1:c.956G>A, NP_000394.2:p.Gly319Glu, NP_001008217.1:p.Gly319Glu, NP_001121093.1:p.Gly319Glu, XM_005245833.1:c.1066G>A, XM_005245834.1:c.1066G>A, XM_005245835.1:c.1066G>A, XM_005245836.1:c.874G>A, XM_005245837.1:c.764G>A, XP_005245890.1:p.Gly356Ser, XP_005245891.1:p.Gly356Ser, XP_005245892.1:p.Gly356Ser, XP_005245893.1:p.Gly292Ser, XP_005245894.1:p.Gly255Glu

rs28940884 [Homo sapiens]

2. AAGGGCCACATTCGACCTTAAGGA[A/G]GCTGAAAGAACAGTGTGGCTGCCGG
Chromosome: 1:23796722
Gene: GALE (GeneView)
Functional Consequence: missense
Allele Origin: G(germline)/A(germline)
Clinical significance: Pathogenic
Validated: by 1000G, by cluster, by frequency
Global MAF: C=0.0082/41
HGVS: NC_000001.10:g.24123212T>C, NC_000001.11:g.23796722T>C, NG_007068.1:g.9083A>G, NM_000403.3:c.770A>G, NM_001008216.1:c.770A>G, NM_001127621.1:c.770A>G, NP_000394.2:n.Val297Asn, NP_001008217.1:n.Val257Asn

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

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

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

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

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

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

Description: UDP-galactose-4-epimerase
Transcript (Including UTRs):
Position: chr1:23,994,676-23,999,884 **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 AGLB4 Search

Display Settings: Summary, 20 per page, Sorted by SNP_ID

Search results: Items: 1 to 20 **221431**

1. rs17378497 has merged into rs355206 [Homo sapiens]

ATATAAACTGCTTTTCAGTTAAAGG[A/C]AAACCTGCTTTTGACCCCTTCCAAA

Chromosome: 1:49206502
Gene: AGLB4 (GeneView)
Functional Consequence: intron variant
Validated: by 1000G, by 2hit 2allele, by cluster, by frequency
Global MAF: A=0.4091/2049
HGVS: NC_000001.10:g.49672174C>A, NC_000001.11:g.49206502C>A, NM_001323573.1:c.413+39268G>T, NM_001323574.1:c.413+39268G>T, NM_001323575.1:c.377+39268G>T, NM_032785.3:c.377+39268G>T, NR_136623.1:n.410+39268G>T, XM_005271284.1:c.413+39268G>T, XM_011542308.2:c.413+39268G>T, XM_011542310.2:c.413+39268G>T, XM_017002595.1:c.377+39268G>T, XM_017002596.1:c.377+39268G>T, XM_017002597.1:c.377+39268G>T, XM_017002598.1:c.377+39268G>T

2. rs17379875 has merged into rs4926831 [Homo sapiens]

TGCTAAGCTCTGTGTTGAGCCTCA[C/T]GATTTGTCAAATGATGAGATATGG

Chromosome: 1:49824429
Gene: AGLB4 (GeneView)
Functional Consequence: intron variant
Validated: by 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap
Global MAF: T=0.2883/1444
HGVS: NC_000001.10:g.50290101C>T, NC_000001.11:g.49824429C>T, NM_001323573.1:c.157+26967G>A, NM_001323574.1:c.157+26967G>A, NM_001323575.1:c.157+26967G>A, NM_032785.3:c.157+26967G>A, NR_136623.1:n.315+26967G>A, XM_005271284.1:c.157+26967G>A, XM_011542308.2:c.157+26967G>A, XM_011542310.2:c.157+26967G>A, XM_017002595.1:c.157+26967G>A, XM_017002596.1:c.157+26967G>A, XM_017002597.1:c.157+26967G>A, XM_017002598.1:c.157+26967G>A

Search details: AGLB4[All Fields]

Recent activity: AGLB4 (221431) SNP, Titin (3871) SNP, Titin TTN (1) SNP, dystrophin DMD (6) SNP, dystrophin (7) SNP

SNP GALE Search

Display Settings: Summary, 20 per page, Sorted by SNP_ID

Search results: Items: 1 to 20 **2843**

1. rs28940885 [Homo sapiens]

CCCAGCTGGCCCAAGAGGAGCTGG[A/G]GTGGACAGCAGCCTTAGGGCTGGAC

Chromosome: 1:23796183
Gene: GALE (GeneView)
Functional Consequence: missense
Allele Origin: G(germline)/A(germline)
Clinical significance: Pathogenic
Validated: by 1000G, by cluster, by frequency
Global MAF: T=0.0020/10
HGVS: NC_000001.10:g.24122673C>T, NC_000001.11:g.23796183C>T, NG_007068.1:g.9622G>A, NM_000403.3:c.956G>A, NM_001008216.1:c.956G>A, NM_001127621.1:c.956G>A, NP_000394.2:p.Gly319Glu, NP_001008217.1:p.Gly319Glu, NP_001121093.1:p.Gly319Glu, XM_005245833.1:c.1066G>A, XM_005245834.1:c.874G>A, XM_005245835.1:c.1066G>A, XM_005245836.1:c.874G>A, XM_005245837.1:c.764G>A, XP_005245890.1:p.Gly356Ser, XP_005245891.1:p.Gly356Ser, XP_005245892.1:p.Gly356Ser, XP_005245893.1:p.Gly292Ser, XP_005245894.1:p.Gly255Glu

2. rs28940884 [Homo sapiens]

AAGGGCCACATTCGACCTTAAGGA[A/G]GCTGAAAGAACAGTGTGGCTGCCGG

Chromosome: 1:23796722
Gene: GALE (GeneView)
Functional Consequence: missense
Allele Origin: G(germline)/A(germline)
Clinical significance: Pathogenic
Validated: by 1000G, by cluster, by frequency
Global MAF: C=0.0082/41
HGVS: NC_000001.10:g.24123212T>C, NC_000001.11:g.23796722T>C, NG_007068.1:g.9083A>G, NM_000403.3:c.770A>G, NM_001008216.1:c.770A>G, NM_001127621.1:c.770A>G, NP_000394.2:p.Ile297Asn, NP_001008217.1:n.Ile257Asn

Search details: GALE[All Fields]

Recent activity: GALE (2843) SNP, AGLB4 (221431) SNP, Titin (3871) SNP, Titin TTN (1) SNP, dystrophin DMD (6) SNP

Базы данных, содержащие информацию по белкам.

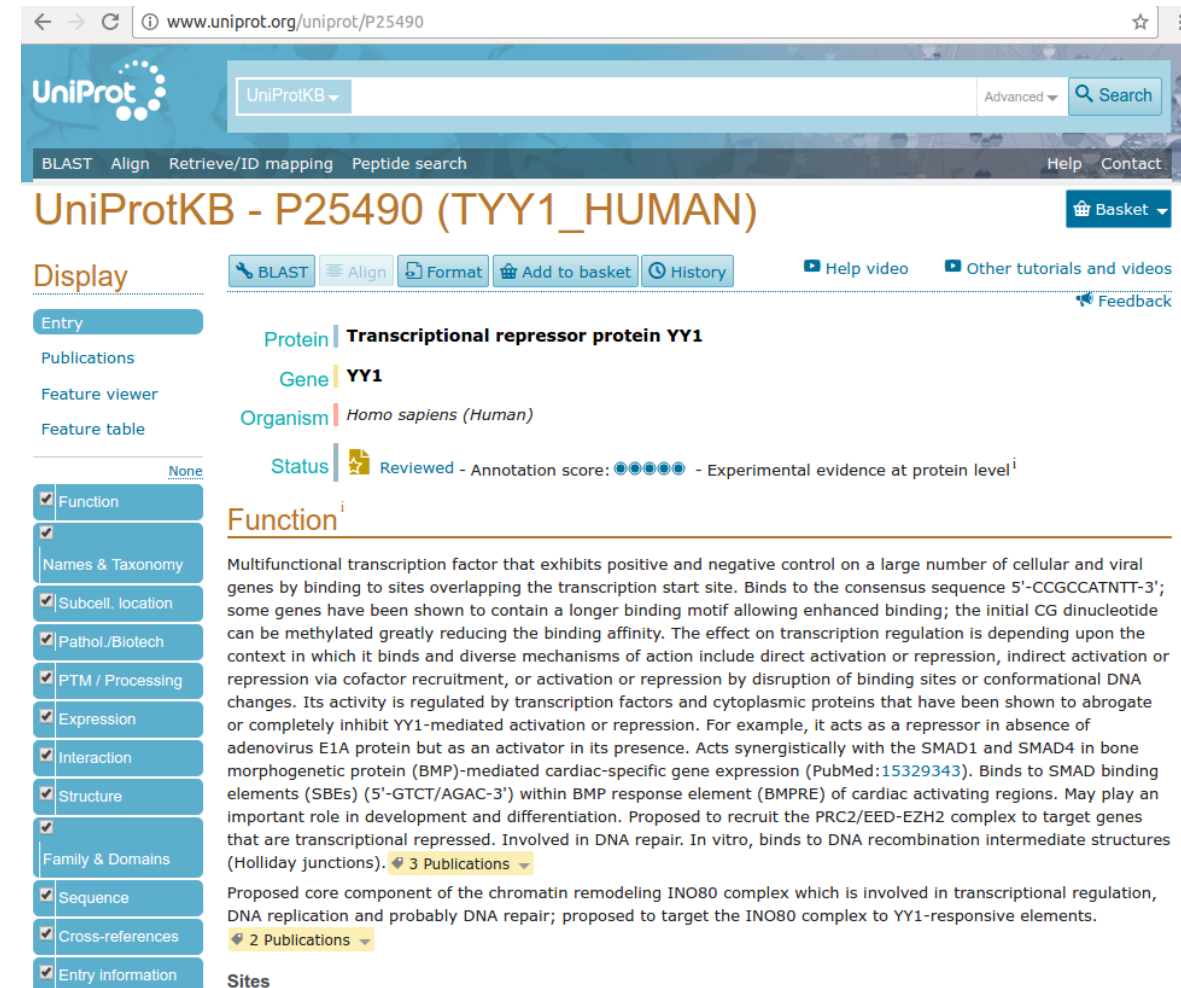


- UniProt
- NCBI Protein
- PDB
- И другие

Uniprot (Universal Protein Resource)

- Доступна через интернет по адресу <http://www.uniprot.org/>
- Содержит информацию по структуре и функциям белков 457010 бактерий, 168308 вирусов, 12163 архей, 894013 эукариот.
- База содержит два раздела: 1) Swiss-Prot, включающий карточки белков, проаннотированные вручную экспертами; 2) TrEMBL, включающий карточки белков, сформированные автоматическими средствами.
- Идентификаторы двух типов 1) шесть знаков: цифры и заглавные буквы 2) короткое название белка_организм. Например, белок YY1 человека P25490 и TYY1_HUMAN, мыши - Q00899 и TYY1_MOUSE.
- Содержит информацию по
 - ✓ синонимам и функциям
 - ✓ Локализации в клетке
 - ✓ Пост-трансляционным модификациям
 - ✓ Экспрессии
 - ✓ Взаимодействиям белка
 - ✓ Последовательности и др.

Информационная карточка белка.



UniProtKB - P25490 (TYY1_HUMAN)

Display

Entry
Publications
Feature viewer
Feature table

Protein | **Transcriptional repressor protein YY1**
Gene | **YY1**
Organism | *Homo sapiens (Human)*
Status | Reviewed - Annotation score: ●●●●●● - Experimental evidence at protein levelⁱ

Functionⁱ

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

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.

Sites

NCBI Protein

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

Информационная карточка белка.

NCBI Protein Search

GenBank: AAH65366.1

YY1 transcription factor [Homo sapiens]

GenBank: AAH65366.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to: [v]

LOCUS AAH65366 414 aa linear PRI 15-JUL-2006

DEFINITION YY1 transcription factor [Homo sapiens].

ACCESSION AAH65366

VERSION AAH65366.1

DBSOURCE accession [BC065366.1](#)

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (residues 1 to 414)

AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,D., Jones,J., Moore,T., May,S.T., Wang,J., Hayes,F.

[Region](#)
/db_xref="CDD:296200"
385..407
/region_name="C2H2 Zn finger"
/note="C2H2 Zn finger [structural motif]"
/db_xref="CDD:275368"
order(385,390,403,407)
/site_type="other"
/note="Zn binding site [ion binding]"
/db_xref="CDD:275368"

[Site](#)
1..414
/gene="YY1"
/gene_synonym="DELTA"
/gene_synonym="NF-E1"
/gene_synonym="UCRBP"
/gene_synonym="YIN-YANG-1"
/coded_by="BC065366.1:46..1290"
/db_xref="GeneID:7528"
/db_xref="HGNC:HGNC:12856"
/db_xref="MIM:600013"

[CDS](#)
ORIGIN
1 masgdtlyia tdgsempaei velheiev et ipvetiETT VgeeeeeeDD dedggggdHg
61 gggghgghag hhhhhhhhhh pPmialqplv tddptqvhhh qeVilvqtre evvggddsdg
121 lraedgfedq ilipvpapag gdddyieqtl vtvaagksg gggssssggg rvkkgggkks
181 gkksylsgga gaaggggadp gnkkweqkqv qiktlegefs vtmwssdekk didhetvvee
241 qiigensppd yseymtgkkl ppggipgidl sdpkqlaefa rmkprkiked daprtiacph
301 kgctkmfrdn samrkhhlth gprvhvcaec gkafvesskl krhqlvhtge kpfcqctfegc
361 gkrfsldfnl rthvrihtgd rpyvcpdfgc nkkfaqstnl kshilthaka knnq

[Protein 3D Structure](#)
Crystal Structure of Mbd1 Yy1 Complex
PDB: 4C5I
Source: Homo sapiens

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

[MODBASE, Database of Comparative Protein Structures](#) [MODBASE, Database of Comparat...]

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

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

[reagents](#) [ExactAntigen/Labome]

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

[Recent activity](#)
Turn Off Clear

YY1 transcription factor [Homo sapiens] Protein

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

PDB - Protein Data Bank

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

Информационная карточка белка.

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

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 127823 Biological Macromolecular Structures

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

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

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

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

Biological Assembly 1

1DGB
HUMAN ERYTHROCYTE CATALASE
DOI: 10.2210/pdb1dgb/pdb

Classification: **OXIDOREDUCTASE**
Deposited: 1999-11-23 Released: 2000-02-11
Deposition author(s): [Putnam, C.D.](#), [Arvai, A.S.](#), [Bourne, Y.](#), [Tainer, J.A.](#)
Organism: [Homo sapiens](#)

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

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.2 Å
R-Value Free: 0.227
R-Value Work: 0.172

wwPDB Validation

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

Literature

Download Primary Citation

Active and inhibited human catalase structures: ligand and NADPH binding and catalytic mechanism.
[Putnam, C.D.](#), [Arvai, A.S.](#), [Bourne, Y.](#), [Tainer, J.A.](#)
(2000) J.Mol.Biol. **296**: 295-309
PubMed: [10656833](#) [Search on PubMed](#)

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

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



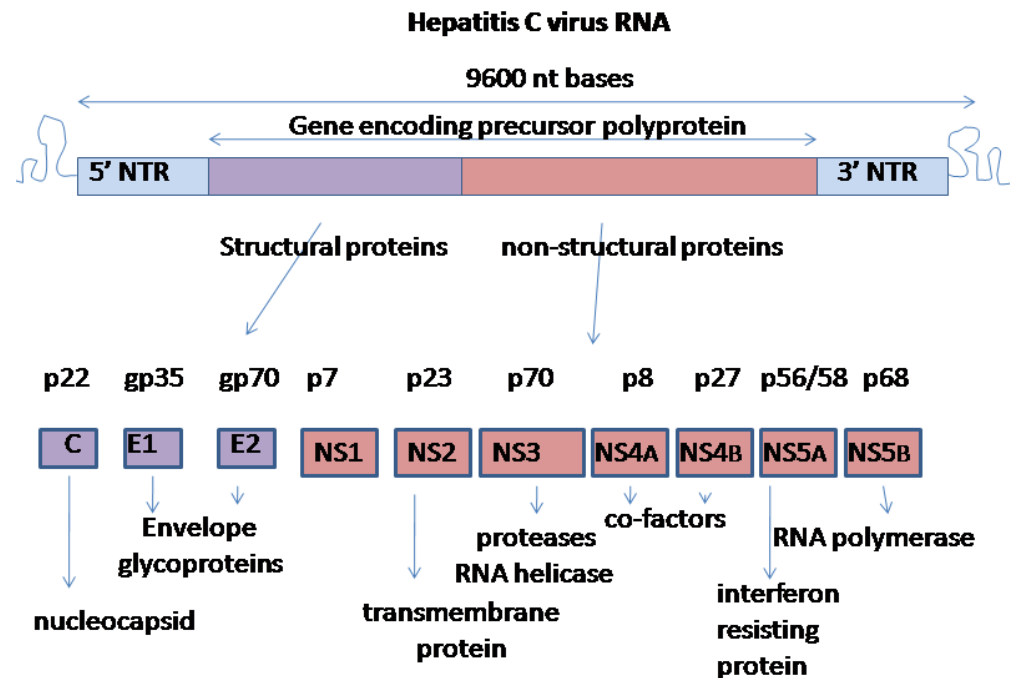
Информационная карточка UniProtKB - P27958 (POLG_HCVH)

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

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

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

Информационная карточка UniProtKB - P27958 (POLG_HCVH)



Protein names

Genome polyprotein
 Cleaved into the following 11 chains:

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

Organism¹ Hepatitis C virus genotype 1a (isolate H) (HCV)

Базы данных, содержащие информацию по микроРНК.



- miRBase
- miRDB
- miRTarBase
- И другие

miRBase

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

Информационная карточка микроРНК.

www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0003716


miRBase MANCHESTER 1824

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

Stem-loop sequence mmu-mir-302b

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

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




MicroRNA and Target Gene Description:

miRNA Name	hsa-miR-302b-3p	miRNA Sequence	UAAGUGCUUCAUGUUUUAGUAG
Previous Name	hsa-miR-302b		
Target Score	100	Seed Location	984, 1137, 1538, 1600, 3494
NCBI Gene ID	55432	GenBank Accession	NM_018566
Gene Symbol	YOD1	3' UTR Length	5171
Gene Description	YOD1 deubiquitinase		

3' UTR Sequence

```

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



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

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

miRTarBase

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

Информационная карточка микроРНК.

miRTarBase

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

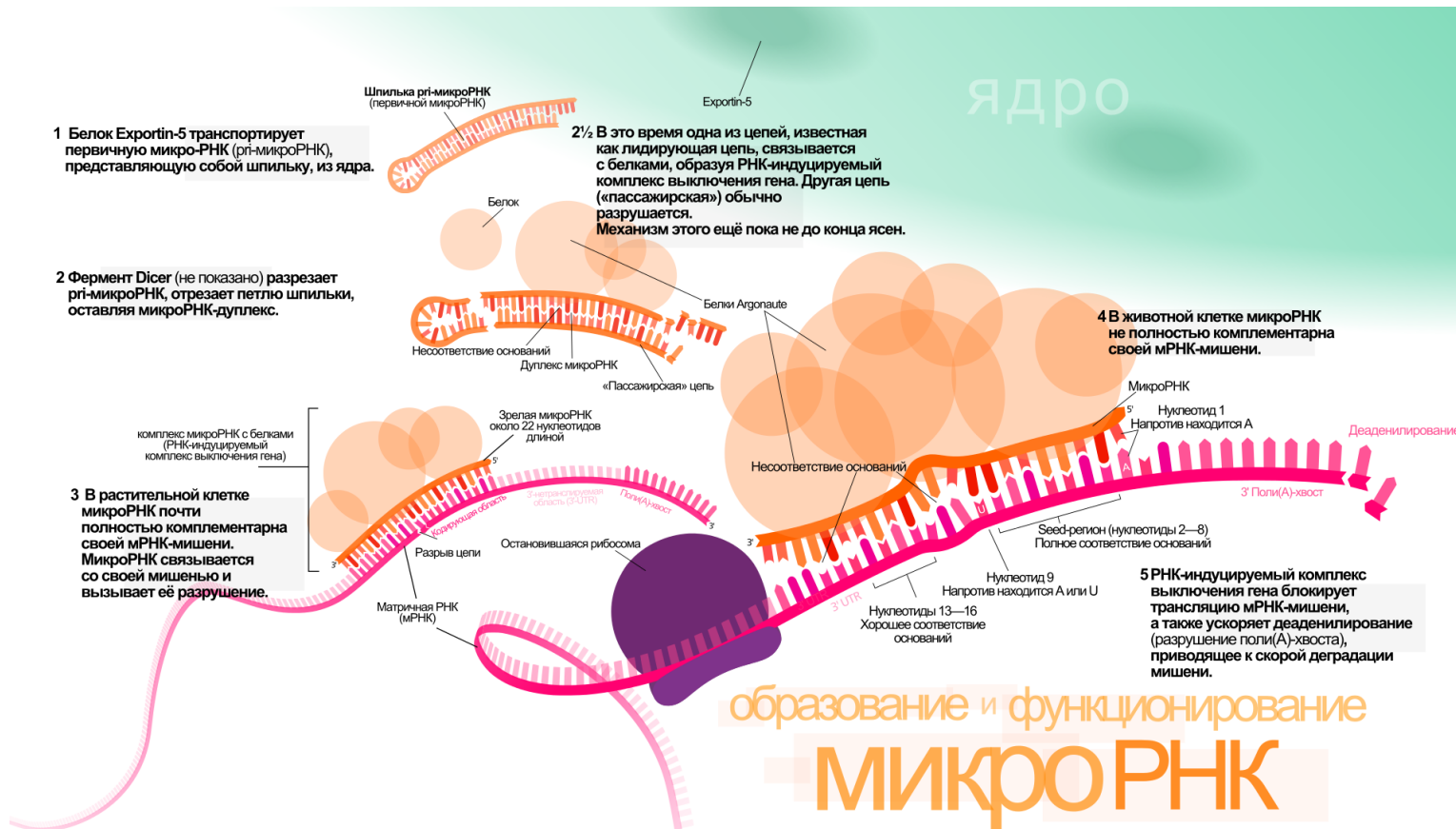
miRNA Target Gene Evidences Expression TCGA Gene Set Enrichments Network ERROR Report

pre-miRNA Information

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

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

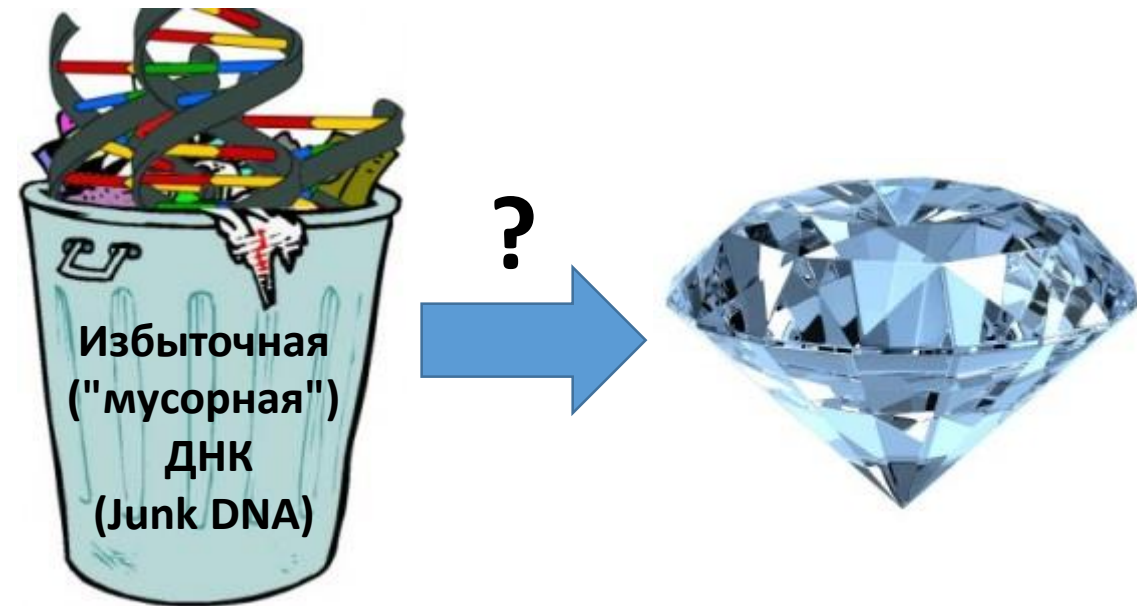
А какие еще классы некодирующих РНК Вы можете назвать?



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

А какие еще классы некодирующих РНК Вы можете назвать?

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



Базы данных, содержащие информацию по метаболитам.



ChEBI

PubChem



- ChEBI
- PubChem
- HMDB
- И другие

ChEBI


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

Информационная карточка метаболита.

← → ↻ <https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:15347> ☆

This website uses cookies. By continuing to browse this site, you are agreeing to the use of our site cookies. To find out more, see our Terms of Use. OK

EMBL-EBI Services Research Training About us


 ChEBI Search
Examples: iron*, InChI=1S/H2O/h1H2, water Advanced

Home | Advanced Search | Browse | Documentation | Download | Tools | About ChEBI Preferences Submit
Contact us

ChEBI > Main

CHEBI:15347 - acetone

Main | ChEBI Ontology | Automatic Xrefs | Reactions | Pathways | Models

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

- [Find compounds which contain this structure](#)
- [Find compounds which resemble this structure](#)
- [Take structure to the Advanced Search](#)

[more structures >>](#)

Wikipedia i License i

Acetone (systematically named propanone) is the [organic compound](#) with the [formula](#) (CH₃)₂CO. It is a colorless, volatile, flammable liquid, and is the simplest [ketone](#). Acetone is [miscible](#) with [water](#) and serves as an important [solvent](#) in its own right, typically for cleaning purposes in the laboratory. About 6.7 million [tonnes](#) were produced worldwide in 2010, mainly for use as a solvent and production of [methyl methacrylate](#) and [bisphenol A](#). It is a common building block in [organic chemistry](#). Familiar

PubChem

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

Информационная карточка метаболита.

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

NIH NLM National Center for Biotechnology Information

PubChem OPEN CHEMISTRY DATABASE Search Compounds

Compound Summary for CID 180 Download Share Help

Acetone

► Cite this Record

STRUCTURE VENDORS PHARMACOLOGY LITERATURE PATENTS BIOACTIVITIES

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

Acetone is a colorless liquid used as a solvent and an antiseptic. It is one of the ketone bodies produced during ketoacidosis. ► from MeSH

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

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

HMDB

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

Информационная карточка метаболита.

Showing metabocard for Acetone (HMDB01659)

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

Show Metabolites with Similar Structures

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

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

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

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

Что это за соединение?

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

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

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



Базы данных, содержащие информацию по заболеваниям.



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



OMIM

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

Информационная карточка заболевания.

125853 ICD+

DIABETES MELLITUS, NONINSULIN-DEPENDENT; NIDDM

Alternative titles; symbols

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

Other entities represented in this entry:

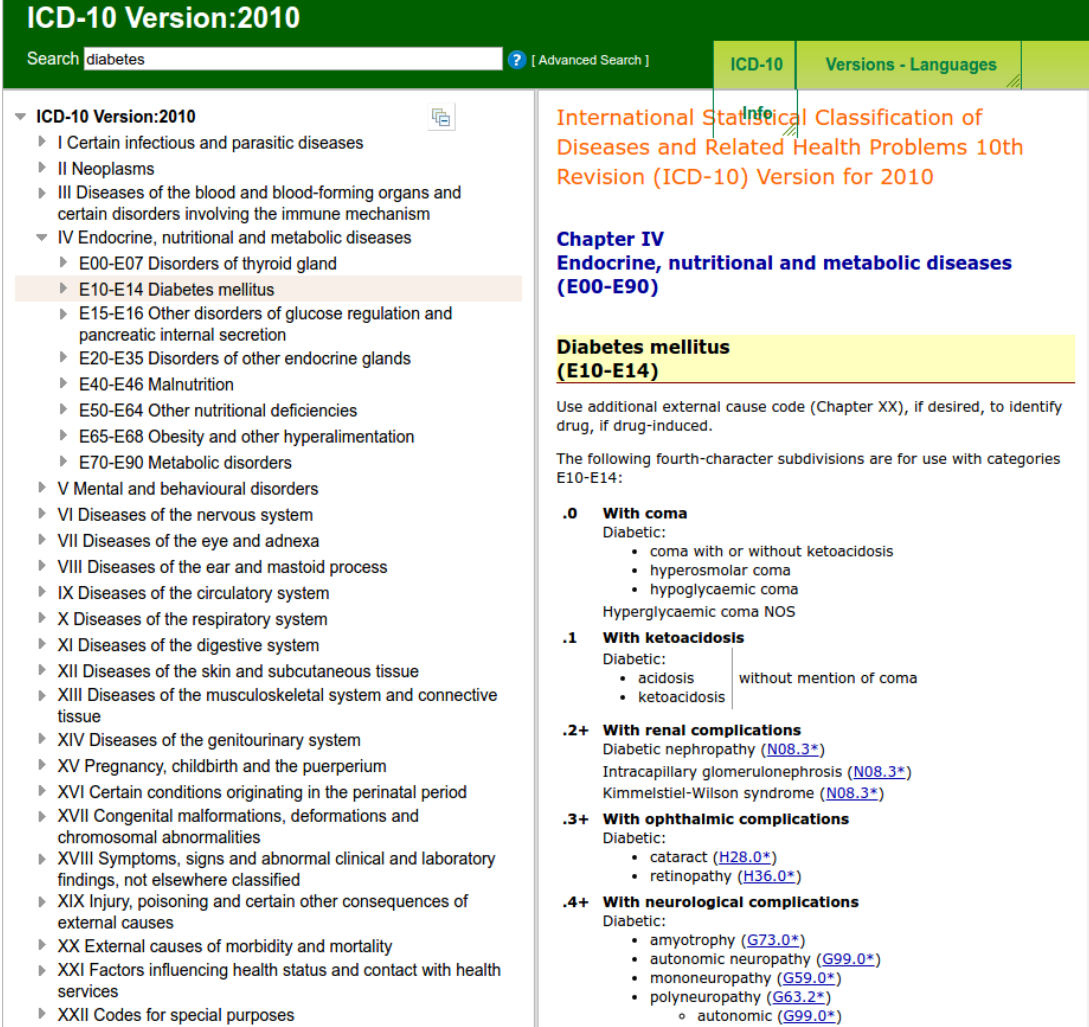
INSULIN RESISTANCE, SUSCEPTIBILITY TO, INCLUDED
DIABETES MELLITUS, TYPE 2, PROTECTION AGAINST, INCLUDED

Phenotype-Gene Relationships

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

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

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



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

ICD-10 Version:2010

Search [Advanced Search]

ICD-10 Versions - Languages

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

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

**Diabetes mellitus
(E10-E14)**

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

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

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

Disease Ontology

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

Информационная карточка заболевания.

The screenshot shows the Disease Ontology website interface. At the top, there is a navigation menu with links for Home, Tutorial, Resources, Downloads, FAQ, About, and Contact Us. Below the menu is the Disease Ontology logo. A search bar contains the word "diabetes" and a "Go" button. To the right of the search bar is an "Advanced Search" link. Below the search bar is a "Navigation" panel with a tree view of disease categories. The "Diabetes Mellitus" term is selected and highlighted. To the right of the navigation panel is a "Metadata" panel with a "Visualize" button. The metadata panel displays the following information:

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

At the bottom of the metadata panel, there is a link to "Add an item to the term tracker".

Оцените, сколько в настоящее время известно заболеваний?



Оцените, сколько в настоящее время известно заболеваний?

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



Базы данных, содержащие информацию по биологическим путям/процессам.



- Gene Ontology
- KEGG
- Reactome
- И другие

Gene Ontology

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

Информационная карточка биологического процесса.

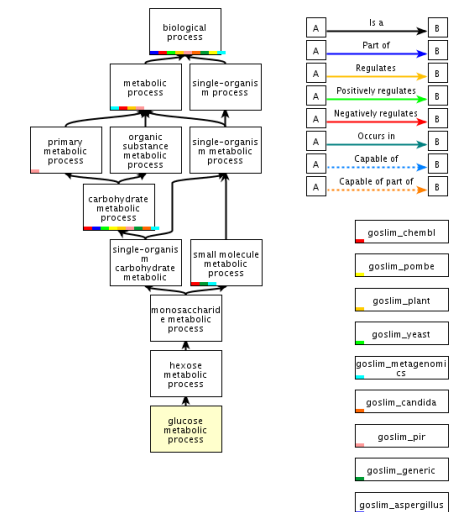
glucose metabolic process

Term Information

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

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

View this term in QuickGO.



KEGG - Kyoto Encyclopedia of Genes and Genomes

- Доступна через интернет по адресу <http://www.genome.jp/kegg/>
- Содержит описание биологических процессов на молекулярно-генетическом уровне, составленное вручную экспертами.
- Идентификаторы: трехбуквенный идентификатор организма и пятизначное число. Например, Glycolysis/Gluconeogenesis имеет идентификатор hsa00010.
- Содержит информацию по
 - ✓ Описанию
 - ✓ Входящим в процесс генам/белкам, метаболитам и др.
 - ✓ Ассоциированным 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, glyceraldehyde-3P, glyceraldehyde-3P, phosphoenolpyruvate, and pyruvate [MD:M00001]. Acetyl-CoA, another important precursor metabolite, is produced by oxidative decarboxylation of pyruvate [MD:M00307]. When the enzyme genes of this pathway are examined in completely sequenced genomes, the reaction steps of three-carbon compounds from glyceraldehyde-3P to pyruvate form a conserved core module [MD:M00002], which is found in almost all organisms and which sometimes contains operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway of glucose from noncarbohydrate precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths [MD:M00003].	
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.

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

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

Below the table, there is a section titled 'Locations in the PathwayBrowser'. The main part of the screenshot is a detailed pathway diagram showing various metabolic steps and molecules. Key sub-pathways highlighted include 'Digestion of dietary carbohydrate', 'Galactose catabolism', 'Fructose metabolism', 'Pentose phosphate pathway (hexose monophosphate shunt)', and 'Catabolism of glucuronate to xylulose-5-phosphate'. A summary box at the bottom of the diagram states: 'These pathways together are responsible for: 1) the extraction of energy and carbon skeletons for biosyntheses from dietary sugars and related molecules; 2) the short-term storage of glucose in the body (as glycogen) and its mobilization and synthesis of glucose from pyruvate during extended fasts.'

breakdown is a major source and conversion to glucose yield pyruvate, glycogen can be synthesized from p

inner mitochondrial mem

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)	Protein Database
BioSystems	Genome	PubChem BioAssay
Bookshelf	Genome Reference Consortium (GRC)	PubChem Compound
ClinicalTrials.gov	HIV-1	PubChem Substance
ClinVar	HomoloGene	PubMed
CloneDB (formerly Clone Registry)	Influenza Virus	PubMed Central (PMC)
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	RefSeqGene
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
Database of Short Genetic Variations (dbSNP)	NCBI Help Manual	Third Party Annotation (TPA) Database
GenBank	NCBI Pathogen Detection Project	Trace Archive
Gene	NCBI Website Search	UniGene
Gene Expression Omnibus (GEO) Database	Nucleotide Database	UniGene Library Browser
Gene Expression Omnibus (GEO) Datasets	Online Mendelian Inheritance in Man (OMIM)	Viral Genomes
Gene Expression Omnibus (GEO) Profiles	PopSet	Virus Variation

Результат поискового запроса к ресурсу NCBI по ключевому слову УУ1

Search NCBI databases

[Help](#)

Search

Results found in 33 databases for "yy1"

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

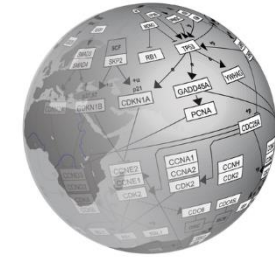
Conserved Domains	7	conserved protein domains
Protein	979	protein sequences
Protein Clusters	1	sequence similarity-based protein clusters
Structure	19	experimentally-determined biomolecular structures

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.

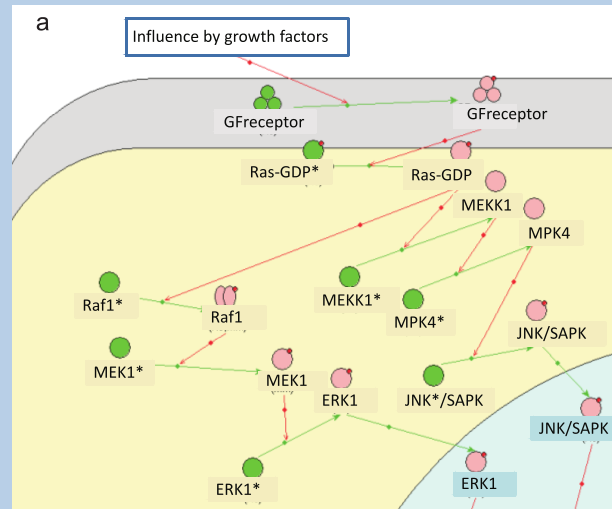


WIKIPATHWAYS
Pathways for the People

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



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



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

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

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

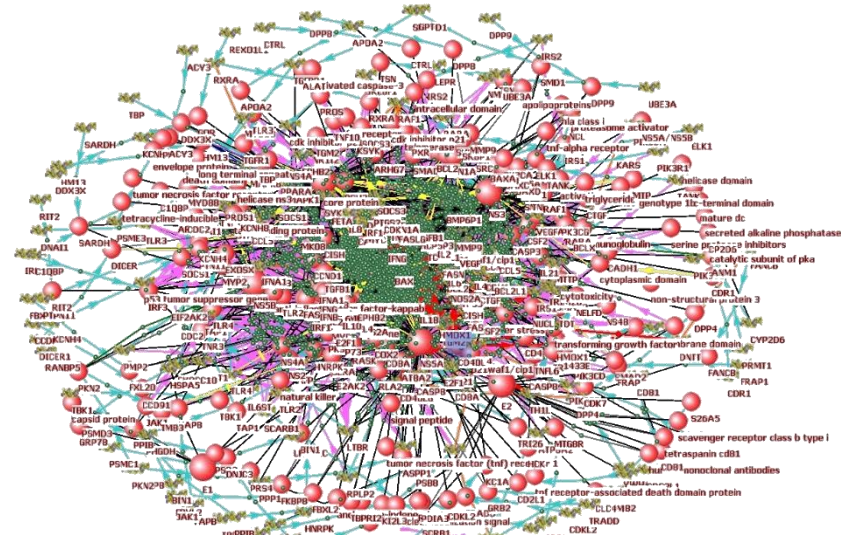
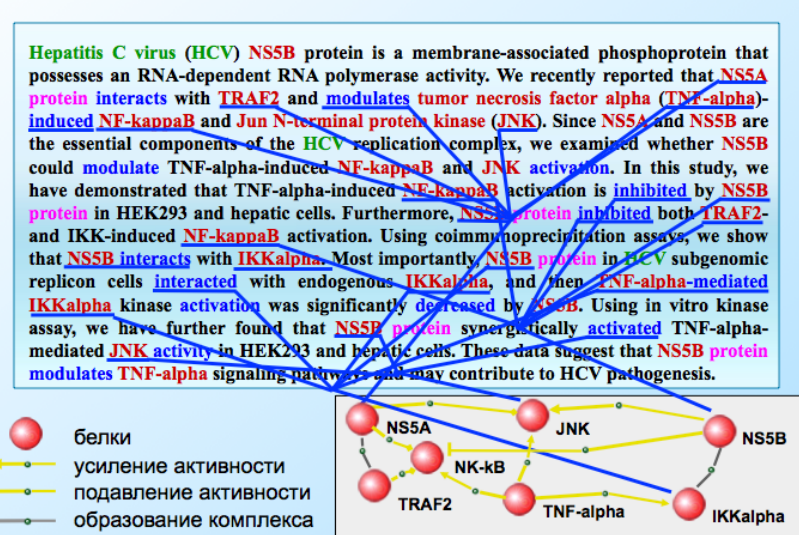
В ИЦиГ СО РАН разработана компьютерная технология автоматического извлечения знаний из текстов научных публикаций и международных патентов ANDSystem. Проведен автоматический анализ более 25 млн. научных публикаций и 10 млн. международных патентов.

Система ANDSystem включает

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

Создана база знаний, содержащая более 15 млн. фактов,

- значимых для биомедицины, включающих взаимосвязи между
- молекулярно-генетическими системами и процессами
- заболеваниями и фенотипическими признаками
- факторами окружающей среды и др.



Pathway Commons



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

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

Pathway Commons интегрирует информацию из 24 баз данных

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

Текстовое поле для ввода названия интересующего гена

Добавление в сеть данных по изменениям, наблюдающимся при раке

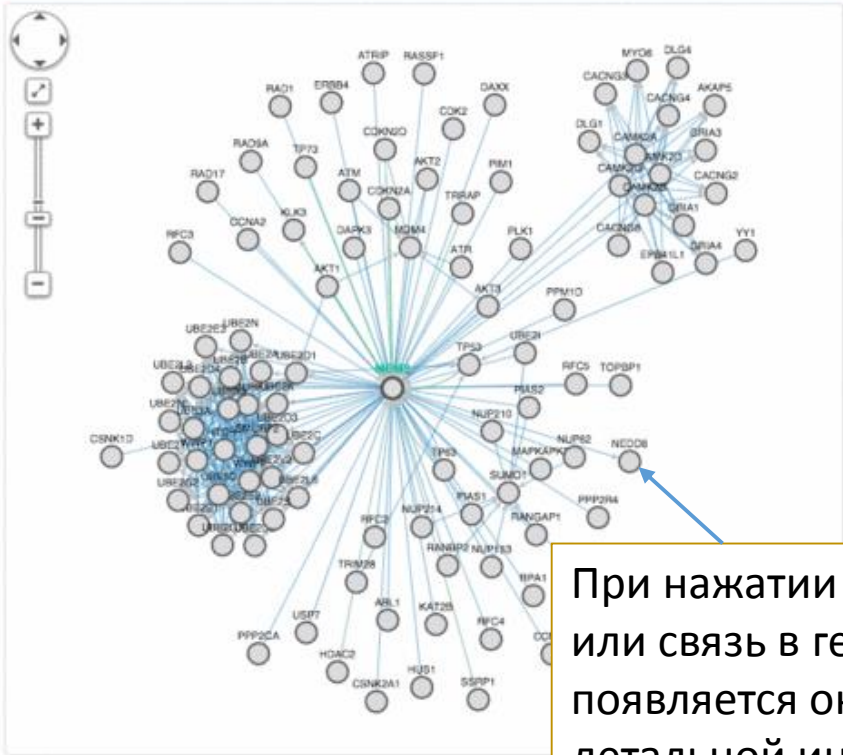
Настройки для фильтрации сети по связям и вершинам

PCViz Pathway Commons Network Visualizer

Genes of interest

MDM2 +

Details Settings Context



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

Details Settings Context

Interaction types

- 362 controls state change x
- 409 controls expression x

Number of genes (468)

Slide left to decrease the number of genes

Query type

Neighborhood

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

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

Details Settings Context

CDT1

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

Aliases: DUP, R152

Description: chromatin licensing and DNA replication factor 1

Chromosome Location: 16q24.3

UniProt ID: [Q9H211](#)

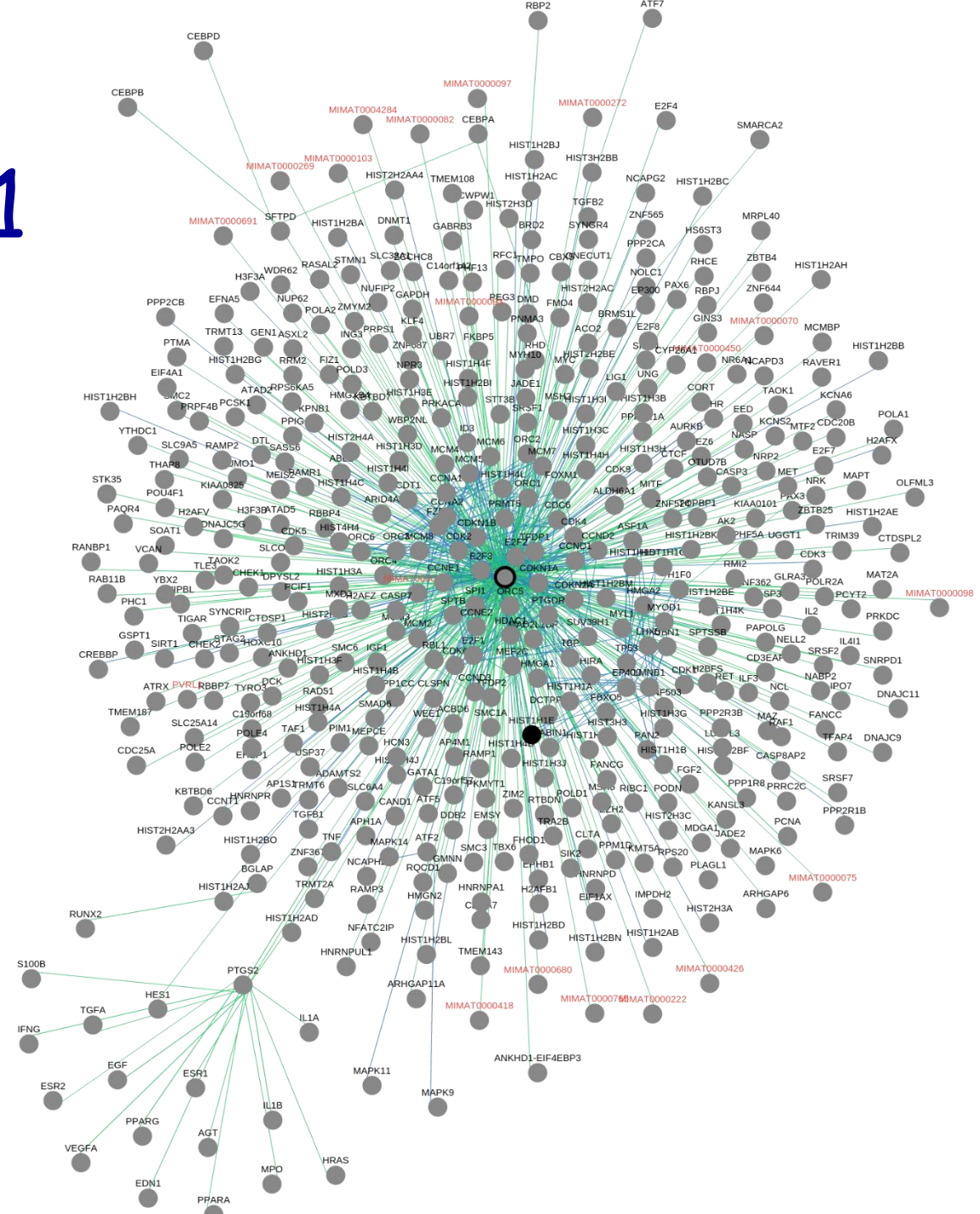
Gene ID: [81620](#)

Добавление гена в список интересующих вас генов

Пример генной сети для гена ретинобластомы RB1

Пример загрузки генной сети в формате SIF

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



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

ANDSystem

STRING

GeneMania

Pathway Commons

Системная биология и моделирование
ОРИГИНАЛЬНЫЕ ИССЛЕДОВАНИЯ

Валеологический журнал генетики и селекции. 2016;20(6):840-847
DOI:10.18699/VJ16.200

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

О.В. Саик¹, Н.А. Козлова², П.С. Демеников¹, Н.В. Иванисенко³, Т.В. Иванисенко³, Д.Е. Исакоздук⁴, М.Н. Понимаренко⁵, О.С. Козлова⁶, Ю.А. Подколякин², И.Н. Лаврик^{1,4}, Н.А. Колчанов¹, В.А. Иванисенко³

¹Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия;
²Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный медицинский университет» Министерства здравоохранения Российской Федерации, Томск, Россия;
³Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный

Saik et al. *BMC Medical Genomics* 2018, 11(Suppl 1):15
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^{1*}, Pavel S. Demenikov¹, Timofey V. Ivanisenko¹, Elena Yu Bragina², Maxim B. Freidin³, Irina A. Goncharova⁴, Victor E. Dosenko⁵, Olga I. Zolotareva¹, Ralf Hofstaedt⁶, Inna N. Lavrik⁶, Evgeny I. Rogaez^{1,7,8,9} and Vladimir A. Ivanisenko¹

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 755 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, TopGene) 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, TLR4, 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)

SCIENTIFIC REPORTS

OPEN

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

Juthatip Manissorn, Supaporn Khamchun, Arada Vinaiphath & 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 α -tubulin, which was decreased, as one of central nodes of protein-protein interactions. Overexpression of α -tubulin (pcDNA6.2-TUBA1A) was then performed and its efficacy was confirmed. pcDNA6.2-TUBA1A could maintain levels of α -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 α -enolase) on apical membrane were dramatically reduced to basal levels by pcDNA6.2-TUBA1A. These findings implicate that α -tubulin has protective roles in kidney stone disease by preventing cell death and cell-crystal adhesion, but on the other hand, enhancing cell proliferation and tissue repair function.

Until now, kidney stone disease is still a public health problem in almost all areas around the world. The disease causes substantial suffering and ultimately end-stage renal disease (ESRD). Unfortunately, the disease mechanisms remain poorly understood. Calcium oxalate (CaOx) is the major chemical component found in clinical stones¹. This type of the stones can be originated from supersaturation of calcium and oxalate ions, leading to crystallization inside renal tubular fluid or urine². CaOx crystals can then nucleate to form "stone nidus" and adhere directly onto apical surface of renal tubular epithelial cells^{3,4}. 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^{5,6}. CaOx crystals also evoke inflammatory processes that can lead to fibrosis, loss of nephron and eventually ESRD⁷.

Even with the aforementioned knowledge, molecular mechanisms of the downstream cellular response remain largely unknown. From our previous expression proteomics study⁸, we have identified a number of proteins with altered levels in MDCK renal tubular cells in response to CaOx crystals. Those altered proteins were involved in various biological processes, i.e. ubiquitination pathway, signal transduction, cellular structure, purine biosynthesis, metabolic enzyme, retinol biosynthesis, cellular transportation, protein degradation, RNA metabolism, RNA binding protein, cell surface antigen, nucleic acid metabolism, antioxidant enzyme, chaperone, carrier protein, and protein biosynthesis. However, functional significance of those altered proteins had not been investigated. In the present study, we thus formed global protein network analysis of those altered proteins. Subsequently, overexpression of a protein, which was one of the central nodes of such protein-protein interactions network, was performed. Moreover, functional investigations were performed to address functional significance of the central node protein and its associated partners in kidney stone disease.

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

175 | (6) 28808 | DOI: 10.1038/srep28808

RESEARCH

Open Access



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

Jen-hwa Chu^{1*}, Jaime E. Hart^{2,3}, Divya Chhabra², Eric Garshick^{2,4}, Benjamin A. Raby^{2,5} and Francine Laden^{2,3,6}

Abstract

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

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

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

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

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

Background

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

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

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

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

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



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

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

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

Yasin Şenbabaoğlu^{1*}, Selçuk Onur Sümer¹, Francisco Sánchez-Vega¹, Debra Bemis¹, Giovanni Criello^{1,6}, Nikolaus Schultz², Chris Sander^{1,6*}

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

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

Abstract

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

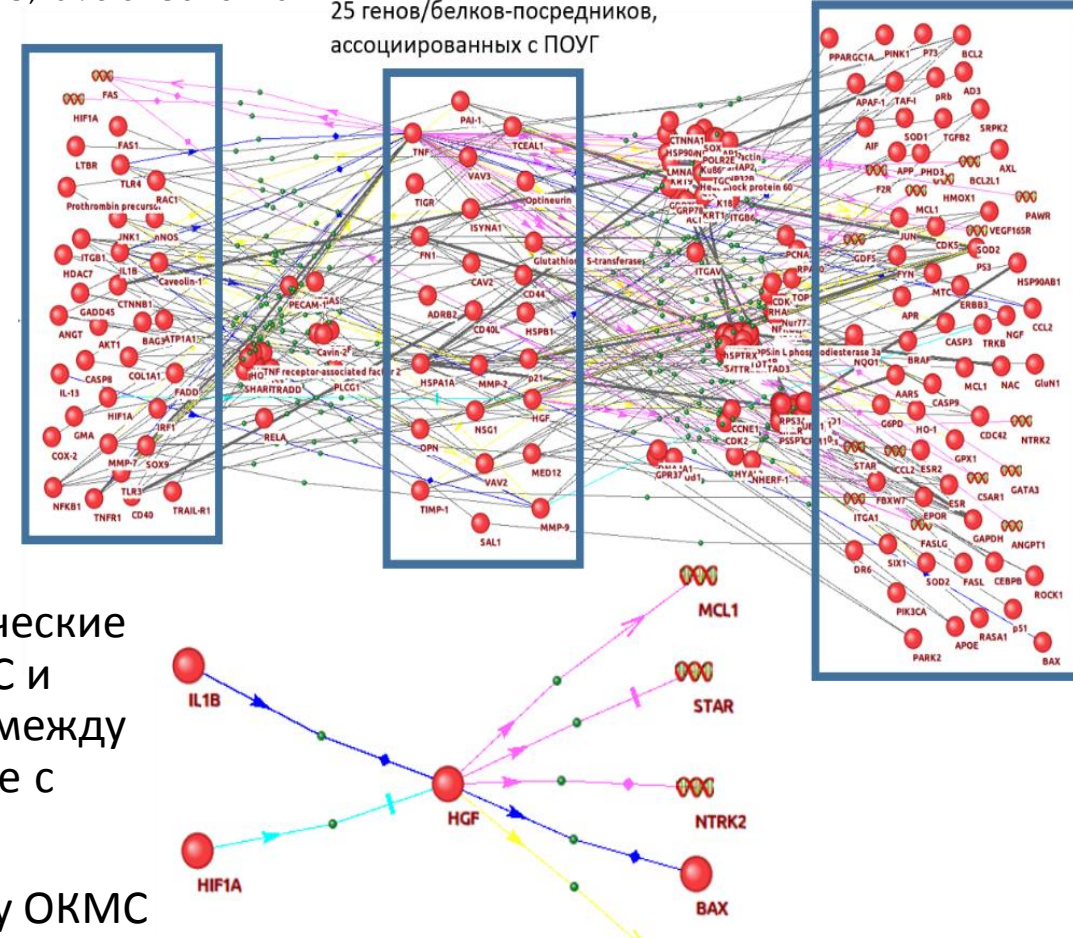
Реконструкция молекулярно-генетических механизмов взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при первичной открыто-угольной глаукоме с помощью системы ANDSystem.

- 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 нарушением функцией генов, ассоциированных с ПОУГ, могут играть важную роль в патогенезе заболевания. Результаты анализа могут быть использованы для поиска новых лекарственных мишеней для терапии ПОУГ.

Белки/гены, вовлеченные в ОКМС

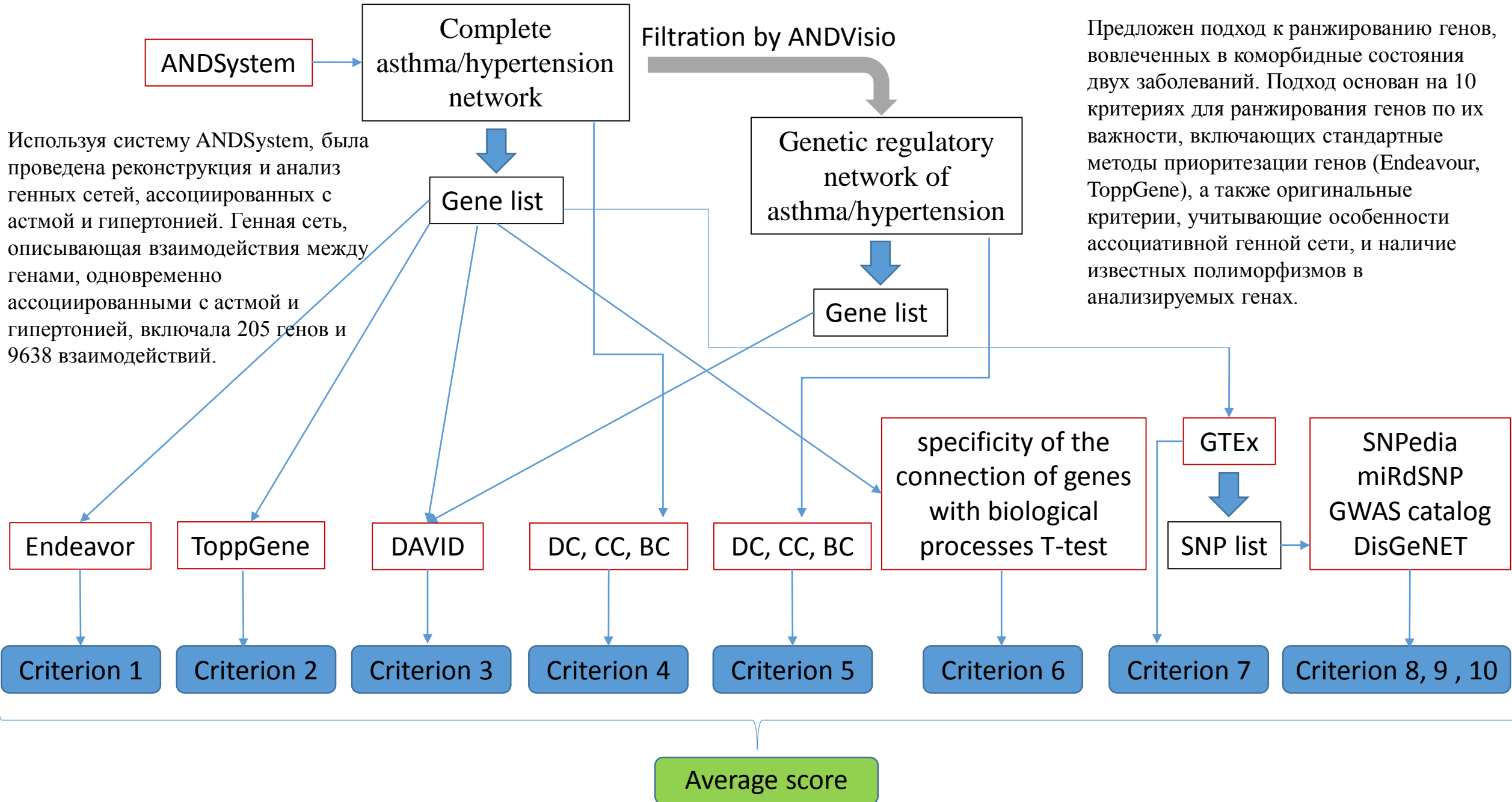
25 генов/белков-посредников, ассоциированных с ПОУГ

Белки/гены, вовлеченные в нейрональный апоптоз

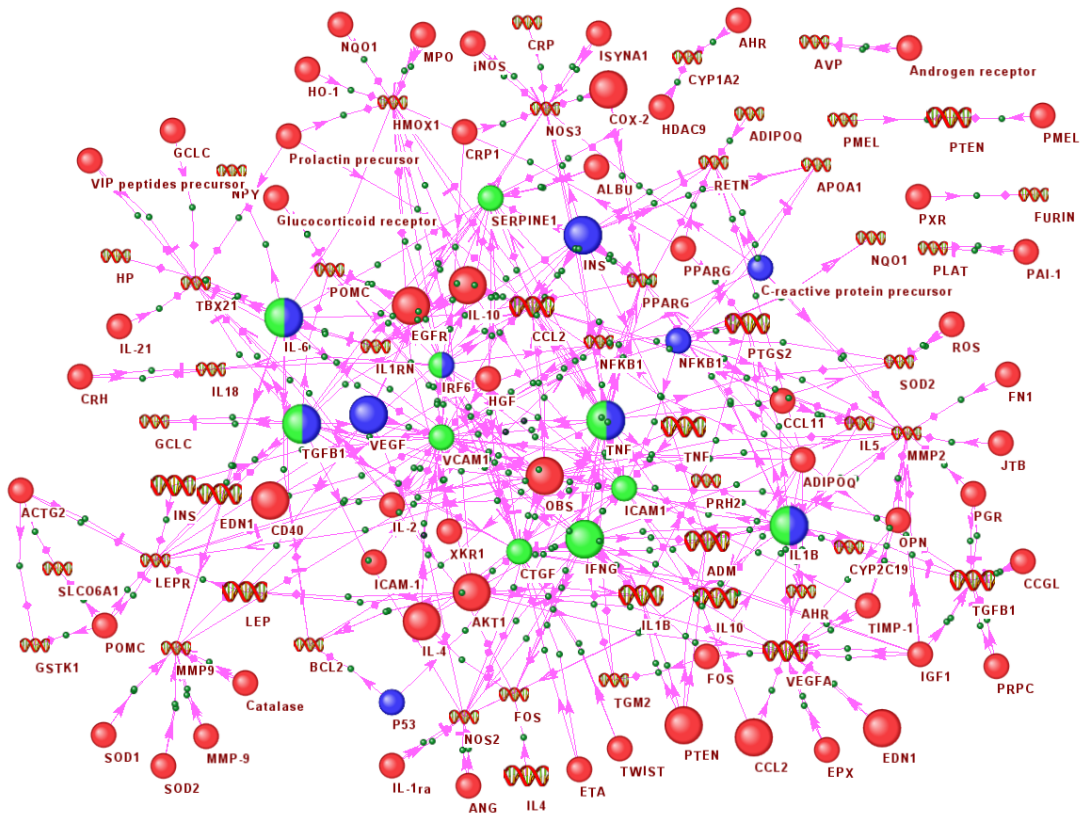


Ассоциативная сеть, описывающая потенциальные взаимодействия между ОКМС и нейрональным апоптозом через молекулярно-генетические цепочки. а) Общая сеть, включающая 25 генов/белков-посредников, ассоциированных с ПОУГ. б) Пример цепочек, проходящих через фактор роста гепатоцитов (HGF), выступающий в качестве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

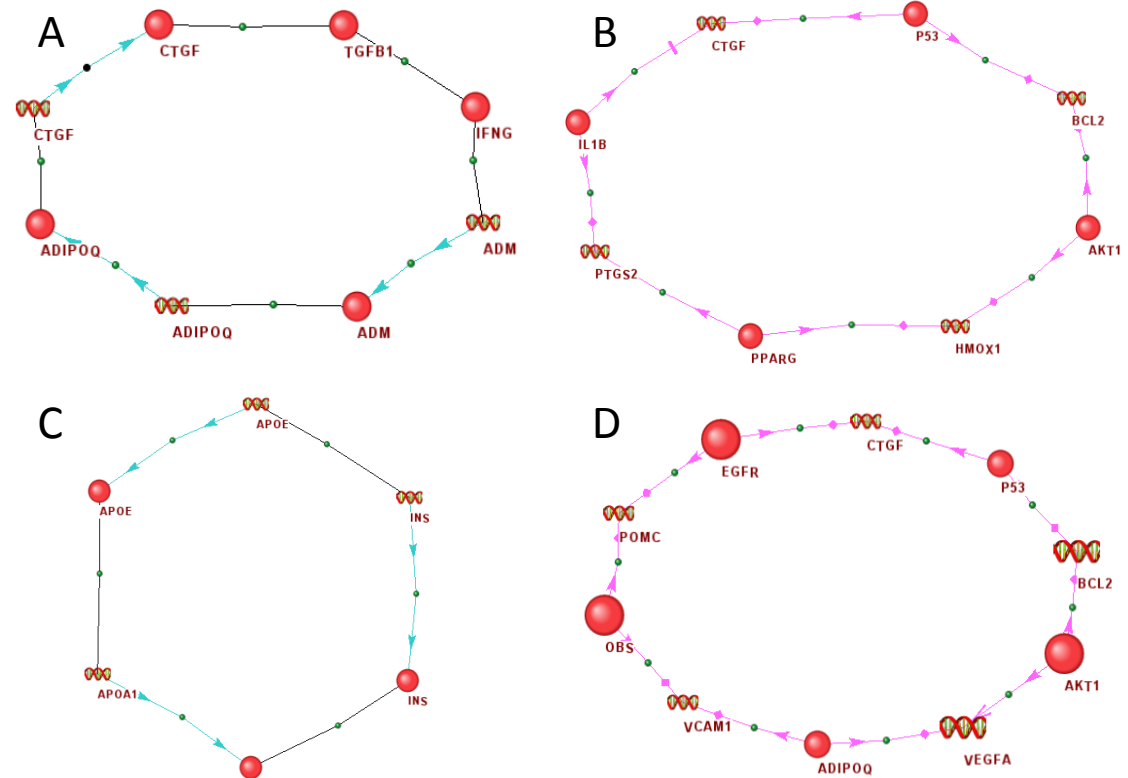
Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks



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



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

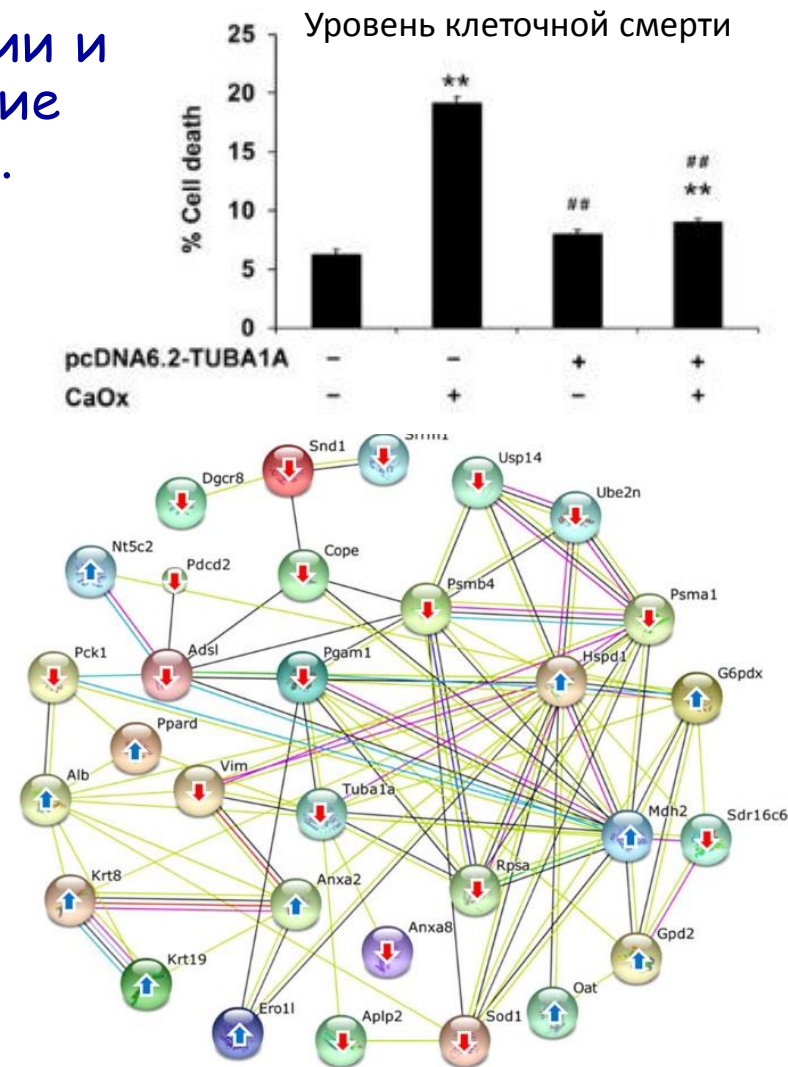
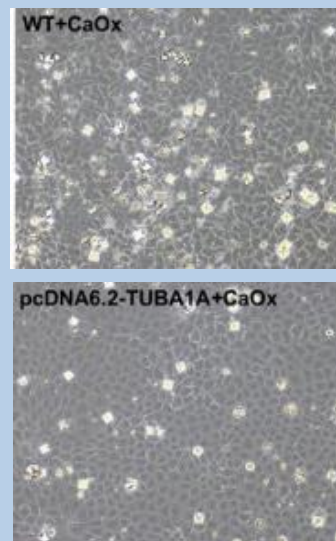


Согласно предложенному подходу гены IL10, TLR4 и CAT имели наивысший приоритет для сочетанного развития астмы и гипертензии

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 α -тубулин, экспрессия которого была снижена в эксперименте, является одним из центральных узлов сети белок-белковых взаимодействий. Для подтверждения ключевой роли α -тубулина была исследована клеточная линия pcDNA6.2-TUBA1A, в которой α -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 α -тубулин играет ключевую роль в патогенезе почечнокаменной болезни.

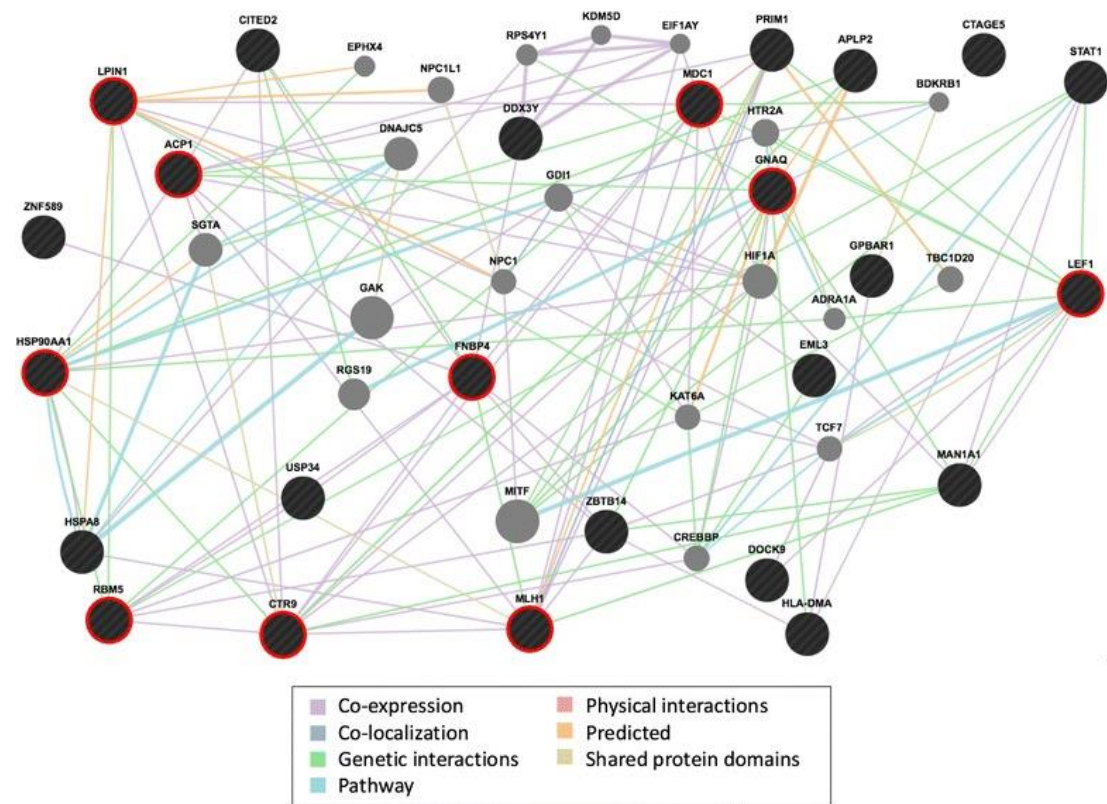
Фазово-контрастная микроскопия почечных трубчатых клеток



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.

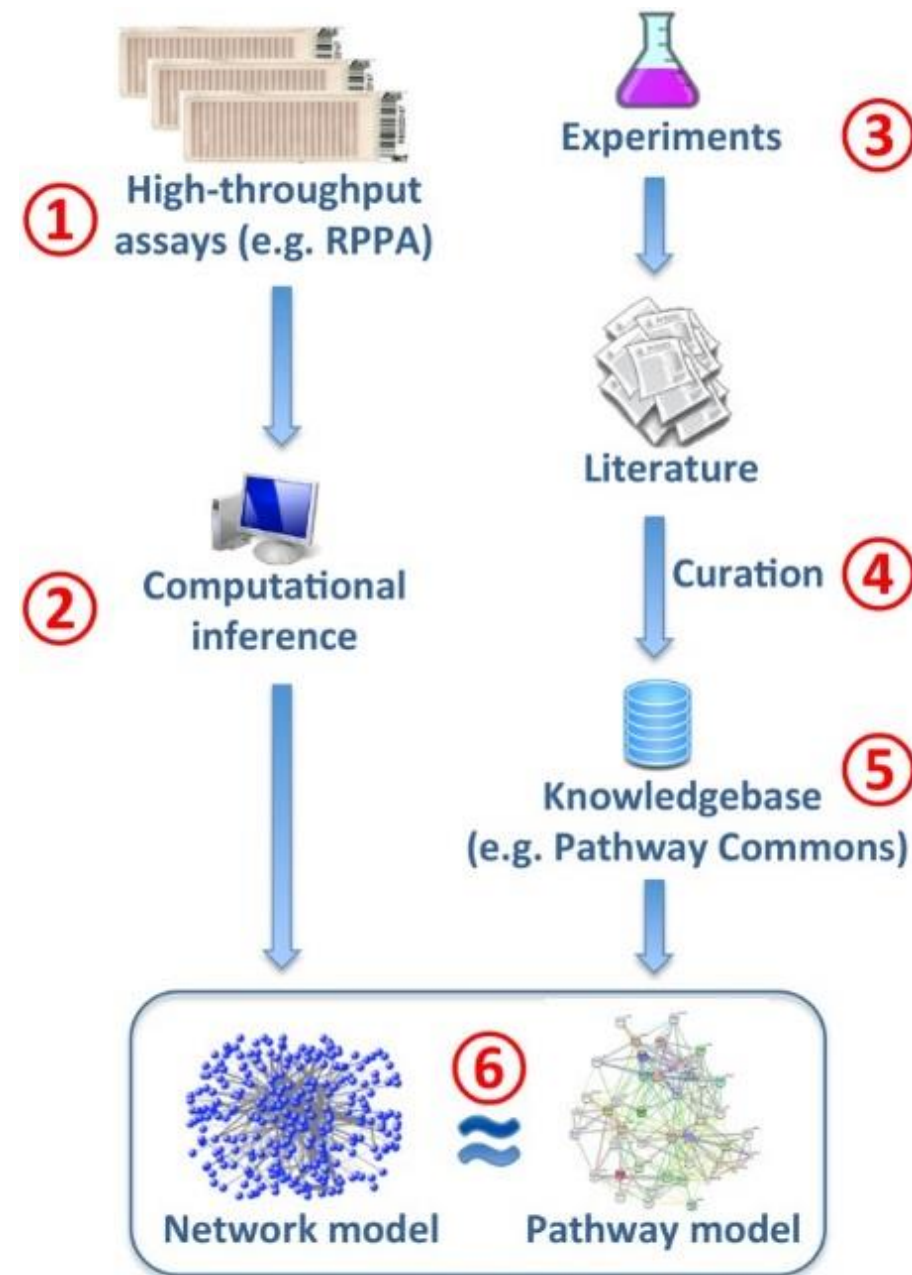
Воздействие загрязненного воздуха связано с неблагоприятными последствиями для здоровья, включая увеличение сердечно-легочной заболеваемости и смертности, а также увеличение риска рака легких. Был проведен полно-геномный анализ уровней экспрессии РНК крови 63 некурящих сотрудников парковочных терминалов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