

Лекция №5

Интеграция разрозненных источников информации по биомедицинской тематике

м.н.с. лаб. компьютерной протеомики
Тийс Евгений Сергеевич,
tiys@bionet.nsc.ru

План лекции

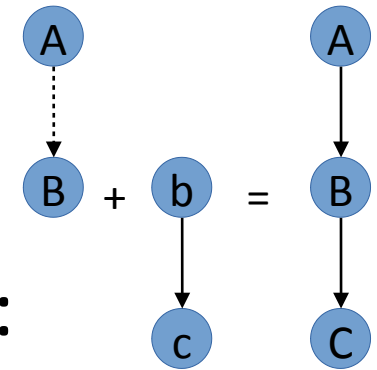
1. Введение.
2. Обзор Интернет-ресурсов, содержащих разнообразную информацию о генных сетях.
3. Ресурсы, интегрирующие биологическую информацию из разнородных источников и представляющие ее в виде генных сетей: ANDSystem, STRING, GeneMania, Pathway Commons.
4. Практическое применение инструментов интеграции.

План лекции

1. Введение.
2. Обзор Интернет-ресурсов, содержащих разнообразную информацию о генных сетях.
3. Ресурсы, интегрирующие биологическую информацию из разнородных источников и представляющие ее в виде генных сетей: ANDSystem, STRING, GeneMania, Pathway Commons.
4. Практическое применение инструментов интеграции.

Цель интеграции биомедицинских данных – понимание и моделирование механизмов биологических процессов

- Преимущества интеграции биологических данных:
- Единый формат – возможность автоматизации обработки и сбор статистики.
- Единый способ графического и текстового представления – упрощение осмысления, отсутствие штрафов на переключения внимания.
- Возможность выполнения поискового запроса ко всем данным
- Установление эквивалентности объектов из различных источников с учетом синонимии, что позволяет избежать дублирования информации
- Выявление противоречий, ошибок, пробелов в информации
- Возможность выявления наиболее достоверной информации

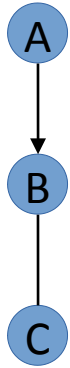


Генная сеть – группа координированно функционирующих генов, взаимодействующих друг с другом как через свои первичные продукты (РНК и белки), так и через разнообразные метаболиты и другие вторичные продукты функционирования генных сетей, которая контролирует какой-либо фенотипический признак организма.

Интегрируемая
информация

Компоненты генной сети:

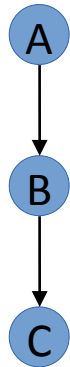
- 1) группа координированно экспрессирующихся генов, составляющая ядро сети;
- 2) белки, кодируемые этими генами;
- 3) низкомолекулярные компоненты (гормоны и другие сигнальные молекулы, энергетические компоненты, метаболиты);
- 4) связи между участниками сети (в том числе отрицательные и положительные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

Подразумеваются взаимодействия,
без посредников из V

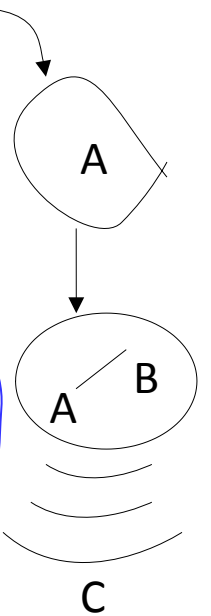


“A gene network is a directed labelled graph, where each node represents a gene and each arc represents a relation between the genes.”

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.



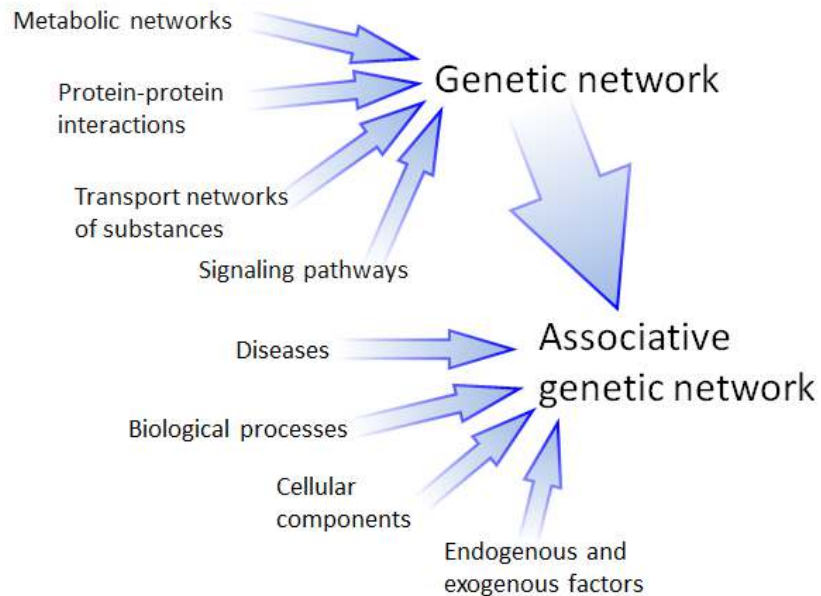
“A gene network is a collection of effective interactions, describing the multiple ways through which one gene affects all the others to which it is connected.”

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.

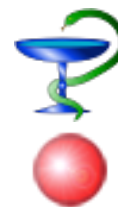
Ассоциативные генные сети являются молекулярно-генетическими сетями, ассоциированными с какими либо биологическими процессами, фенотипическими признаками или заболеваниями.

Ассоциативные генные сети в качестве вершин включают следующие типы объектов:

- (1) **молекулярно-генетические объекты**. К этим типам объектов относятся гены, РНК, белки, метаболиты, клеточные компоненты;
- (2) **биологические процессы и системы** – метаболические пути, пути передачи сигналов, транспортные пути и т.п.);
- (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).

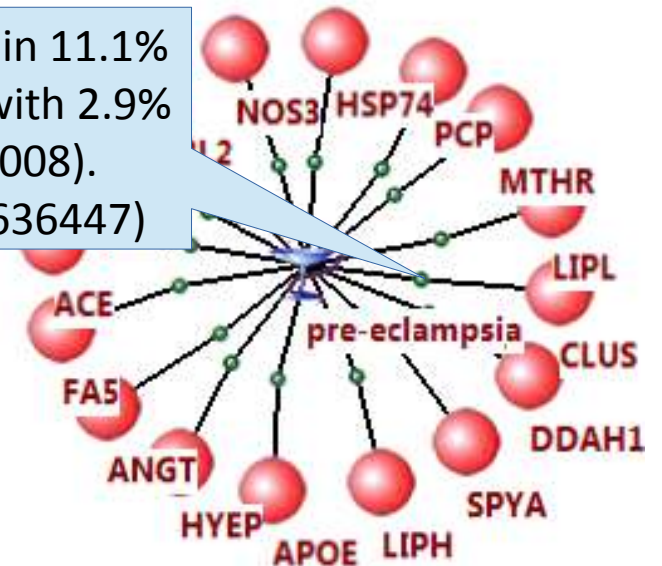


The N291S variant was identified in 11.1% of pre-eclampsics as compared with 2.9% of pregnancy controls ($p = 0.008$). (Hubel CA et al. 1999, PMID: 10636447)



заболевание

белок



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	PubChemBioAssay
Bookshelf	Genome Reference Consortium (GRC)	PubChemCompound
ClinicalTrials.gov	HIV-1	PubChemSubstance
ClinVar	HomoGene	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 YY1

Search NCBI databases

[Help](#)

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

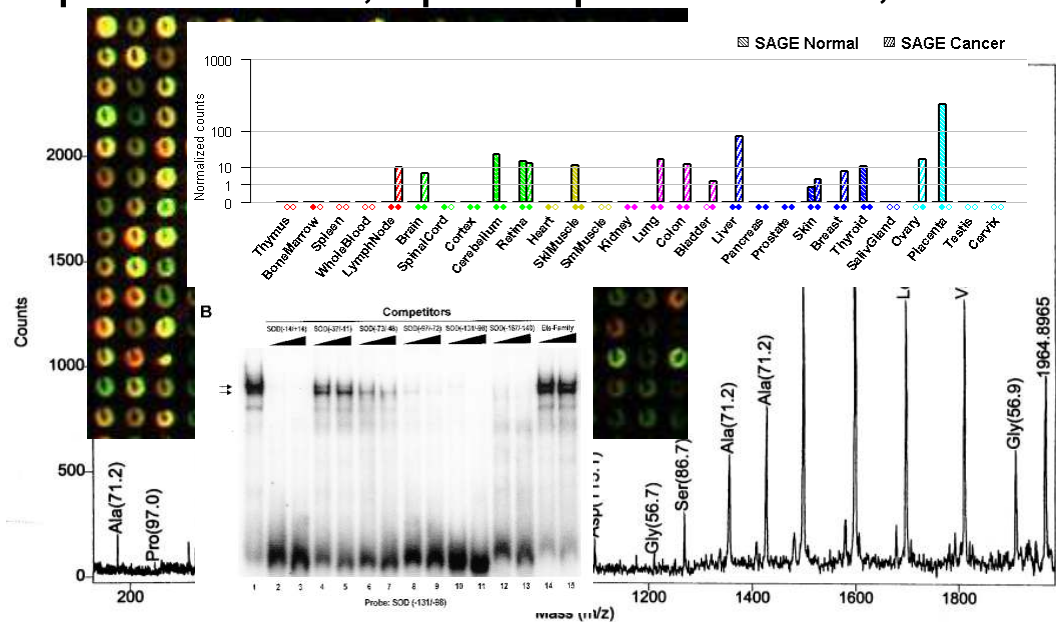
План лекции

1. Введение.
2. Обзор Интернет-ресурсов, содержащих разнообразную информацию о генных сетях.
3. Ресурсы, интегрирующие биологическую информацию из разнородных источников и представляющие ее в виде генных сетей: ANDSystem, STRING, GeneMania, Pathway Commons.
4. Практическое применение инструментов интеграции.

Источники информации

Экспериментальные данные:

Протеомные, транскриптомные, метаболомные



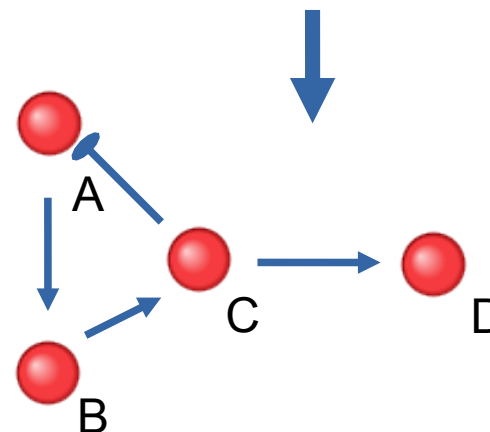
базы данных



PubMed

Скорость появления публикаций **1,4** в минуту

http://www.nlm.nih.gov/bsd/medline_cit_counts_yr_pub.html

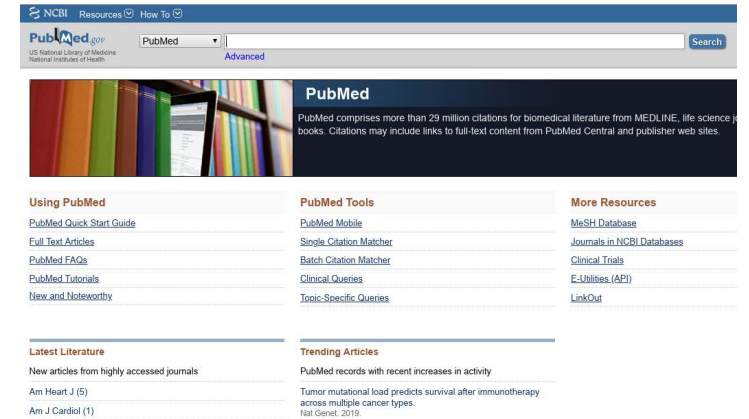
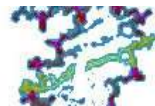


Источники информации

Важнейшим источником информации по биомедицинской тематике являются научные публикации. В системе **PubMed**, которая является крупнейшей базой данных научных статей, доступных через интернет, собрано более 29 миллионов публикаций.

Журнал **Nucleic Acids Research (NAR) database issue** (IF=8,65) публикует списки баз данных. На январь 2019 года работали 1613 баз:

<http://www.oxfordjournals.org/nar/database/c>



Интерфейс PubMed

Категории баз данных

- Nucleotide Sequence Databases
- RNA sequence databases
- Protein sequence databases
- Structure Databases
- Genomics Databases (non-vertebrate)
- Metabolic and Signaling Pathways
- Human and other Vertebrate Genomes
- Human Genes and Diseases
- Microarray Data and other Gene Expression
- Proteomics Resources
- Other Molecular Biology Databases
- Organelle databases
- Plant databases
- Immunological databases
- Cell biology

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:
3950 документа

2. *Panax ginseng* – женьшень обыкновенный:
6911 документов



How To Sign in to NCBI

PubMed Create RSS Create alert Advanced Help

Format: Summary Sort by: Best Match Per page: 20 Send to Filters: [Manage Filters](#)

Search results
Items: 1 to 20 of 3950 << First < Prev Page 1 of 198 Next > Last >>

[Functional characterisation and cell specificity of BvSUT1, the transporter that loads sucrose into the phloem of sugar beet \(*Beta vulgaris* L.\) source leaves.](#)
Nieberl P, Ehrh C, Pommerrenig B, Graus D, Marten I, Jung B, Ludewig F, Koch W, Harms K, Flügge U, Neuhaus HE, Hedrich R, Sauer N. *Plant Biol (Stuttg)*. 2017 May;19(3):315-326. doi: 10.1111/plb.12546. Epub 2017 Feb 1. PMID: 28075052 [Similar articles](#)

Results by year

[Download CSV](#)

How To Sign in to NCBI

PubMed Create RSS Create alert Advanced Help

Format: Summary Sort by: Best Match Per page: 20 Send to Filters: [Manage Filters](#)

Search results
Items: 1 to 20 of 6911 << First < Prev Page 1 of 346 Next > Last >>

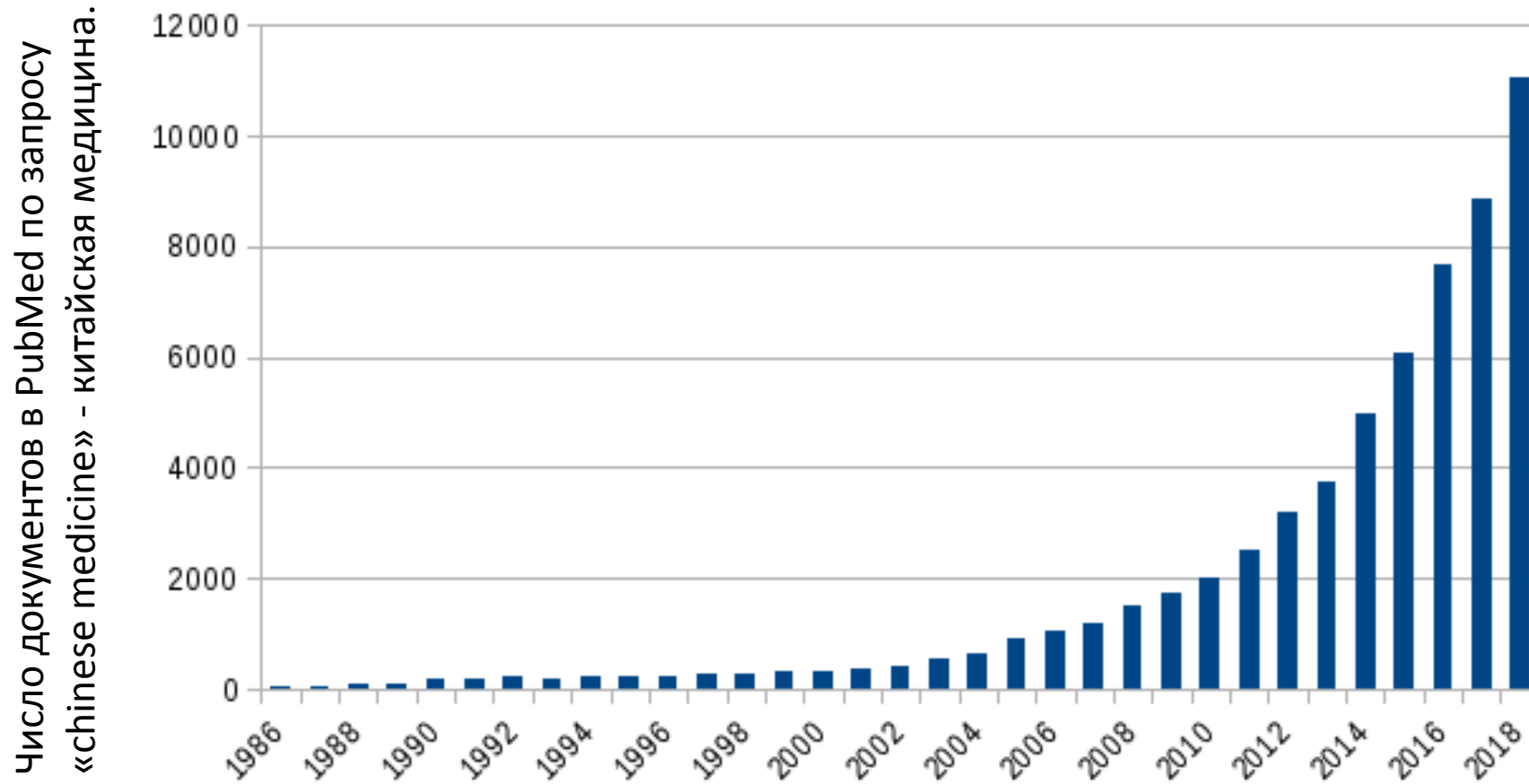
[Panax ginseng and Panax quinquefolius: From pharmacology to toxicology.](#)
1. Mancuso C, Santangelo R. *Food Chem Toxicol*. 2017 Sep;107(Pt A):362-372. doi: 10.1016/j.fct.2017.07.019. Epub 2017 Jul 8. Review. PMID: 28698154 [Similar articles](#)

[Ginseng in Dermatology: A Review.](#)
2. Sabouri-Rad S, Sabouri-Rad S, Sahebkar A, Tayarani-Najaran Z. *Curr Pharm Des*. 2017;23(11):1649-1666. doi: 10.2174/1381612822666161021152322. Review.

Results by year

[Download CSV](#)

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



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 HGNC logo and a search bar are visible. The main content area is titled "Symbol Report: YY1". Below this, a table lists key information:

APPROVED SYMBOL	YY1
APPROVED NAME	YY1 transcription factor
HGNC ID	HGNC:12856
PREVIOUS SYMBOLS & NAMES	-
LOCUS TYPE	gene with protein product
CHROMOSOMAL LOCATION	14q32.2
GENE FAMILY	INO80 complex Zinc fingers C2H2-type
HCOP	Orthology Predictions for YY1

Below the table, there are sections for "External links", "HOMOLOGS", "GENE RESOURCES", "NUCLEOTIDE SEQUENCES", "PROTEIN RESOURCES", and "CLINICAL RESOURCES".

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: [Q1THUMG00000150479](#) 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.
- Идентификаторами генов являются числа. Например, ген транскрипционного фактора YY1 человека имеет идентификатор 7528, мыши – 22632, крысы – 24919, *Arabidopsis thaliana* – 826093.

Поиск

Окно для ввода запросов

Сохранение

Организмы

The screenshot shows the NCBI Gene database search results for the query 'yy1'. 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:

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

Other search results include YY1 ID: 534353 (transcription factor) and YY1 ID: 826093 (zinc finger protein). The search results are displayed in a table with columns for Name/Gene ID, Description, Location, and Aliases. The search results are also displayed in a table with columns for Name/Gene ID, Description, Location, and Aliases. The search results are also displayed in a table with columns for Name/Gene ID, Description, Location, and Aliases.

NCBI Gene - Информационная карточка гена

Надежный | <https://www.ncbi.nlm.nih.gov/gene/7528>

NCBI Resources How To Sign in to NCBI

Gene: Search

Advanced Help

Full Report ▾ Send to: ▾ Hide sidebar >>

YY1 YY1 transcription factor [*Homo sapiens* (human)]

Gene ID: 7528, updated on 12-Mar-2017

Summary

Official Symbol YY1 provided by [HGNC](#)
Official Full Name YY1 transcription factor provided by [HGNC](#)
Primary source [HGNC:HGNC:12856](#)
See related [Ensembl:ENSG00000100811](#) [MIM:600013](#); [Vega:OTTHUMG00000150479](#)
Gene type protein coding
RefSeq status REVIEWED
Organism [Homo sapiens](#)
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as DELTA; NF-E1; UCRBP; INO80S; YIN-YANG-1
Summary YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins. The protein is involved in repressing and activating a diverse number of promoters. YY1 may direct histone deacetylases and histone acetyltransferases to a promoter in order to activate or repress the promoter, thus implicating histone modification in the function of YY1. [provided by RefSeq, Jul 2008]
Orthologs [mouse](#) [all](#)

Genomic context

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

Exon count: 5

Annotation release	Status	Assembly	Chr	Location
108	current	GRCh38.p7	14	NC_000014.9

Genome Browsers

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

Идентификатор

Общая информация

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

Более детальная информация

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

NCBI Nucleotide

- <https://www.ncbi.nlm.nih.gov/nucleotide>
- Коллекция нуклеотидных последовательностей из нескольких источников GenBank, RefSeq, TRA и PDB.
- Идентификаторы обычно начинаются с двух заглавных букв, например: NM_003403.4

Nucleotide

G – gene
M – mRNA
P – protein

Версия
документа

Номер mRNA, не совпадает
с номером гена или белка

The screenshot shows the NCBI Nucleotide search interface. The search bar contains 'yy1' and the results are displayed as a list of 3830 nucleotide sequences. The first result is a synthetic construct from Homo sapiens, and the second is a mRNA sequence from Homo sapiens (NM_003403.4). The interface includes filters for species, molecule types, source databases, genetic compartments, sequence length, release date, and revision date. The right sidebar shows 'Results by taxon' with a list of organisms and 'Recent activity' with a list of recent searches.

NCBI Nucleotide. Карточка последовательности гена.

Особенности нуклеотидной последовательности

NCBI Nucleotide search results for **Homo sapiens YY1 transcription factor (YY1), mRNA** (NCBI Reference Sequence: NM_003403.4).

General Information:

- 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

References:

- 1 (bases 1 to 3159): 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. *J. Pathol.* 238 (5), 651-664 (2016). [PubMed ID: 26800240](#)
- 2 (bases 1 to 3159): Nieborak A and Gorecki A. *FEBS Lett.* 590 (6), 838-847 (2016). [PubMed ID: 26910132](#)

Analysis Tools: Run BLAST, Pick Primers, Highlight Sequence Features, Find in this Sequence, Articles about the YY1 gene, Pathways for the YY1 gene.

```
exon      1384..1542
          /gene="YY1"
          /gene_synonym="DELTA; IN0805; NF-E1; UCRBP; YIN-YANG-1"
          /inference="alignment:Splign:2.0.8"
exon      1543..3159
          /gene="YY1"
          /gene_synonym="DELTA; IN0805; NF-E1; UCRBP; YIN-YANG-1"
          /inference="alignment:Splign:2.0.8"
STS       1729..1852
          /gene="YY1"
          /gene_synonym="DELTA; IN0805; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="RH1618"
          /db_xref="UniSTS:42115"
STS       1954..2640
          /gene="YY1"
          /gene_synonym="DELTA; IN0805; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="YY1_3892"
          /db_xref="UniSTS:462950"
STS       2035..2173
          /gene="YY1"
          /gene_synonym="DELTA; IN0805; 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; IN0805; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="G06170"
          /db_xref="UniSTS:49520"
polyA_site 2714
          /gene="YY1"
          /gene_synonym="DELTA; IN0805; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="RH68412"
          /db_xref="UniSTS:51218"
STS       3001..3100
          /gene="YY1"
          /gene_synonym="DELTA; IN0805; 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; IN0805; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="RH44921"
          /db_xref="UniSTS:36888"
polyA_site 3132
          /gene="YY1"
          /gene_synonym="DELTA; IN0805; NF-E1; UCRBP; YIN-YANG-1"
          /standard_name="RH44921"
          /db_xref="UniSTS:36888"
ORIGIN
1 agggcgaacg ggcgagtgcc agcgaggcgg ggcgggctga ggccagcggc gaagtctcgc
61 gaggccgggc ccgagcagag tgtggcggcg gcggcgagat ctgggctcgg gttgaggagt
121 tggattttgt gtggaaggag gcggaggcgc aggaggaagg gggagcggga gcgccggccc
181 ggaggggcgg aggagggcgc gccaggcgcg gccggttcgg cgagcgagg cgagcggggc
241 agccgagacc agcagcgccc gacgagggcg gacgagggcg ccaccgaggc gaggagggcg
301 ggggaagccc gccgcccggc cggcgcccgc cccctcccc cccgcccggc cccctcccc
361 ccgcccgctc gccgcttccc tccctctgcc tctctctccc acggccggcc gcctctcgcc
421 ccgcccgccc gcagccgagg agccgaggcc gcccgggcgg 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
 - ✓ Описанию, функция
 - ✓ структуре нуклеотидной последовательности
 - ✓ продуктам гена (соответствующим белкам)
 - ✓ Биологические пути, в которых ген участвует
 - ✓ Экспрессии
 - ✓ Ортологам
 - ✓ Полиморфизму
 - ✓ 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". Key identifiers include GCID: GC14P100238 and GIFIS: 64. The page features a navigation menu with categories like "Aliases", "Disorders", "Domains", "Drugs", "Expression", "Function", "Genomics", "Localization", and "Orthologs". Below the navigation, there are advertisements for various biotech companies such as EMD MILIPORE, GenScript, ORIGENE, and Vigene. The main content area is titled "Aliases for YY1 Gene" and lists numerous alternative names for the gene, such as "YY1 Transcription Factor", "INO80 Complex Subunit S", "Delta Transcription Factor", "Yin And Yang 1 Protein", "INO80S", "NF-E1", and "YY-1". It also provides "External Ids for YY1 Gene" from HGNC, Entrez Gene, Ensembl, OMIM, and UniProtKB. A section for "Summaries for YY1 Gene" 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/P число. Например, ген YY1 человека имеет идентификатор ENSG00000100811, один из транскриптов - ENST00000636393, ген YY1 мыши - ENSMUSG00000021264.
- Удобен для проведения сравнительных геномных исследований.
- Содержится информация по:
 - ✓Синонимам, описанию, локализации
 - ✓Последовательности
 - ✓Ортологам и паралогам
 - ✓Биологическим процессам
 - ✓Полиморфизмам
 - ✓Экспрессии
 - ✓Регуляции и др.

Информационная карточка гена

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

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

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

Gene-based displays

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

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Gene: YY1 ENSG00000100811

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

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

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

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

Transcripts [Show transcript table](#)

Summary

Name [YY1](#) (HGNC Symbol)

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

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

RefSeq Overlapping RefSeq Gene ID [7528](#) matches and has similar biotype of protein_coding

Ensembl version ENSG00000100811.11

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

Gene type Known protein coding

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

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

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

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

Drag/Select: 64.50 kb Forward strand

Genes (Comprehensive set...)

YY1-005 > retained intron

YY1-002 > TEC

YY1-004 > protein coding

RP11-638I2.6-001 > lincRNA

UCSC (University of California, Santa Cruz) Genome Browser

Информационная карточка гена

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

The screenshot displays the UCSC Genome Browser interface for the Human Gene YY1 (ENST00000262238.8). The top navigation bar includes links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, View, and Help. The main content area is titled "UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly" and shows a genomic track for chromosome 14 (hg38) with coordinates 100,238,985-100,282,792. The track displays various annotations including RefSeq genes, human mRNAs, and DNase I hypersensitivity peak clusters. Below the tracks, the "Human Gene YY1 (ENST00000262238.8) Description and Page Index" is shown, providing a detailed description of the gene and its function. The description states that YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins, involved in repressing and activating a diverse number of promoters. The page index table lists various database links and annotations.

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				

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 g sites overlapping the transcription start site. Binds to the consensus sequence 5'-CCGCCATNTT-3'; some genes have been shown longer binding motif allowing enhanced binding; the initial CG dinucleotide can be methylated greatly reducing the binding affinity. transcription regulation is depending upon the context in which it binds and diverse mechanisms of action include direct activation

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 exon combination: X82208.2, CY866728.1, F500023221, BNAseq intron: single sample

Transcript (uc003adj.3)
Position: chr22:27748277-27801498

Coding Region (uc003adj.3)
Position: chr22:27748277-27801498

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

GEO - Gene Expression Omnibus

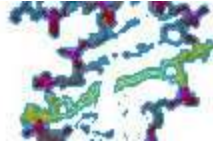
- <https://www.ncbi.nlm.nih.gov/geo/>
- Данные по экспрессии генов. Высокопроизводительные эксперименты по оценке уровней экспрессии генов (методы ДНК-микрочипов и RNA-seq).
- Идентификаторы для DataSet - GDSчисло и для DataSetSeries - GSEчисло. Например, серия экспериментов, посвященных анализу уровня 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

Отношения между платформами, сериями и образцами



Базы данных, содержащие информацию о полиморфизме в генах

dbSNP
Short Genetic Variations



 NCBI

SNPedia



GWAS Catalog

- dbSNP
- ClinVar
- SNPedia
- GWAS

dbSNP

- <https://www.ncbi.nlm.nih.gov/SNP/>
- Содержит информацию по полиморфизму 10 организмов. Для человека представлена информация по более чем 600 млн коротких вариаций.
- Идентификаторы начинаются с двух букв rs далее число. Например, полиморфизм в гене человека YY1 имеет идентификатор rs61992955.
- Содержит информацию по
 - ✓ Частотам встречаемости генетического варианта в разных популяциях
 - ✓ Позиции в геноме
 - ✓ Фланкирующим последовательностям
 - ✓ Эффекту и др.

Информационная карточка вариации

dbSNP Short Genetic Variations

Reference SNP (rs) Report

[← Switch to classic site](#)

rs61992955

Organism	<i>Homo sapiens</i>
Position	chr14:100274488
Alleles	G>A / G>C
Variation Type	SNV Single Nucleo
Frequency	C=0.04944 (6208/ C=0.0509 (1576/3 C=0.030 (150/500

Variant Details | **Genomic**

Clinical Significance	Sequenc
Frequency	GRCh37
Aliases	GRCh37
Submissions	GRCh38
History	GRCh38
	YY1 RefS
	YY1 RefS

Search for rs Search
Example: rs268

[Download](#) [f](#) [t](#) [v](#) [?](#)

Current Build 152
Released October 2, 2018

Clinical Significance Not Reported in ClinVar

Gene : Consequence YY1 : Intron Variant

Publications 0 citations

Genomic View [See rs on genome](#)

Change

NC_000014.8:g.100740825G>C
NC_000014.8:g.100740825G>A
NC_000014.9:g.100274488G>C
NC_000014.9:g.100274488G>A
NG_046908.1:g.40724G>C
NG_046908.1:g.40724G>A

В базе данных dbSNP для гена AGLB4 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 AGLB4 в большей степени подвержен мутагенезу?

Human Gene AGLB4 (uc001cru.2) Description and Page Index

Description: Homo sapiens ATP/GTP binding protein-like 4 (AGLB4), mRNA.
Transcript (Including UTRs):
Position: chr1:48,998,527-50,489,626 **Size:** 1,491,100 **Total Exon Count:** 14 **Strand:** -
Coding Region:
Position: chr1:48,999,845-50,489,468 **Size:** 1,489,624 **Coding Exon Count:** 14

SNP

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

Search results

Items: 1 to 20 of 221431

rs17378497 has merged into rs355206 [Homo sapiens]

1.

ATAAAACTGCTTTTCAGTTAAGG[A/C]AAACCTGCTTTTGACCCCTTTCCAAA
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

rs17379875 has merged into rs4926831 [Homo sapiens]

2.

TGCTAAGCTCTGTTTGAAGCCTCA[C/T]GATTTGTCAAATGAATGAGATATGG
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

Send to:

Find related data

Database:

Search details

AGLB4[All Fields]

Recent activity

AGLB4 (221431)

Titin (3871)

Titin TTN (1)

dystrophin DMD (6)

dystrophin (7)

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

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

SNP

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

Search results

Items: 1 to 20 of 2843

rs28940885 [Homo sapiens]

1.

CCCAGCCTGCCCAAGGAGGCTGG[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_01008216.1: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.

AAGGCCACATTGCGACCTTAAGGA[A/G]GCTGAAAGAACAGTGGCTGCCGG
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, NC_007068.1:g.9083A>G, NM_000403.3:c.770A>G, NM_01008216.1:c.770A>G, NM_001127621.1:c.770A>G, NP_000394.2:p.Lys257Asp, NP_001008217.1:p.Lys257Asp

Send to:

Find related data

Database:

Search details

GALE[All Fields]

Recent activity

GALE (2843)

AGLB4 (221431)

Titin (3871)

Titin TTN (1)

dystrophin DMD (6)

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

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

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

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

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

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

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

SNP

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

Search results: Items: 1 to 20 of 221431

1. rs17378497 has merged into rs355206 [Homo sapiens]

ATATAAACTGCTTTTCAGTTAAGG [A/C] AAACCTGCTTTTGACCCCTTTCCAAA

Chromosome: 1:49206502
Gene: **AGBL4** (GeneView)
Functional Consequence: intron variant
Validated: by 1000G, by 2hit 2allele, by cluster, by frequency
Global MAF: A=0.4091/2049
HGVS: NC_000001.10:g.49672174C>A, NC_000001.11:g.49206502C>A, NM_001323573.1:c.413+39268G>T, NM_001323574.1:c.413+39268G>T, NM_001323575.1:c.377+39268G>T, NM_032785.3:c.377+39268G>T, NR_136623.1:n.410+39268G>T, XM_005271284.1:c.413+39268G>T, XM_011542308.2:c.413+39268G>T, XM_011542310.2:c.413+39268G>T, XM_017002595.1:c.377+39268G>T, XM_017002596.1:c.377+39268G>T, XM_017002597.1:c.377+39268G>T, XM_017002598.1:c.377+39268G>T

2. rs17379875 has merged into rs4926831 [Homo sapiens]

TGCTAAGCTCTGTGTTGAGCCTCA [C/T] GATTTGTCAAATGATGAGATATGG

Chromosome: 1:49824429
Gene: **AGBL4** (GeneView)
Functional Consequence: intron variant
Validated: by 1000G, by 2hit 2allele, by cluster, by frequency, by hapmap
Global MAF: T=0.2883/1444
HGVS: NC_000001.10:g.50290101C>T, NC_000001.11:g.49824429C>T, NM_001323573.1:c.157+26967G>A, NM_001323574.1:c.157+26967G>A, NM_001323575.1:c.157+26967G>A, NM_032785.3:c.157+26967G>A, NR_136623.1:n.315+26967G>A, XM_005271284.1:c.157+26967G>A, XM_011542308.2:c.157+26967G>A, XM_011542310.2:c.157+26967G>A, XM_017002595.1:c.157+26967G>A, XM_017002596.1:c.157+26967G>A, XM_017002597.1:c.157+26967G>A, XM_017002598.1:c.157+26967G>A

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

Search results: Items: 1 to 20 of 2843

1. rs28940885 [Homo sapiens]

CCCAGCCTGCCCAAGGAGGCTGG [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

2. rs28940884 [Homo sapiens]

AAGGCCACATTGCAGCCTTAAGGA [A/G] GCTGAAAGACAGTGGCTGCCGG

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.Lys257Asp, NP_001008217.1:p.Lys257Asp

ClinVar

- <https://www.ncbi.nlm.nih.gov/clinvar/>
- Содержит информацию по генетическим вариантам человека и их связи с заболеваниями.
- Идентификаторы – числа, но также используются более разветнутые названия. Например, вариант гена человека YY1 имеет идентификатор 91950, а также NM_003403.4:c.1115C>G (p.Thr372Arg).
- Содержит информацию по
 - ✓ Клинической значимости генетического варианта
 - ✓ Типу генетического варианта
 - ✓ Позиции в геноме и в белке

Информационная карточка варианта

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

[...more](#)

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 4096

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 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 полиморфизму человека и его 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 аллельных вариантов в разных популяциях
 - ✓ Публикациям и др.

Информационная карточка полиморфизма

Navigation: [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	

Population	(A:A)	(A:C)	(C:C)
CEU	35%	55%	10%
HCB	45%	45%	10%
JPT	55%	35%	10%
YRI	10%	15%	75%
ASW	15%	15%	70%
CHB	45%	45%	10%
CHD	45%	45%	10%
GIH	35%	55%	10%
LWK	15%	15%	70%
MEX	15%	15%	70%
MKK	15%	15%	70%
TSI	35%	55%	10%
AVG	35%	55%	10%

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 search results for rs2766692, including a table of studies and a list of associations.

Examples: breast cancer, rs7329174, Yang, 2q37.1, HBS1L

GWAS / Search / rs2766692

Refine search results

Show results for

- Studies 1
- Associations 1
- Catalog traits 9

Filter results by

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

Odds ratio from to

Beta coefficient

Search results for rs2766692

Download association results

Expand all studies

Studies

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



Associations

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

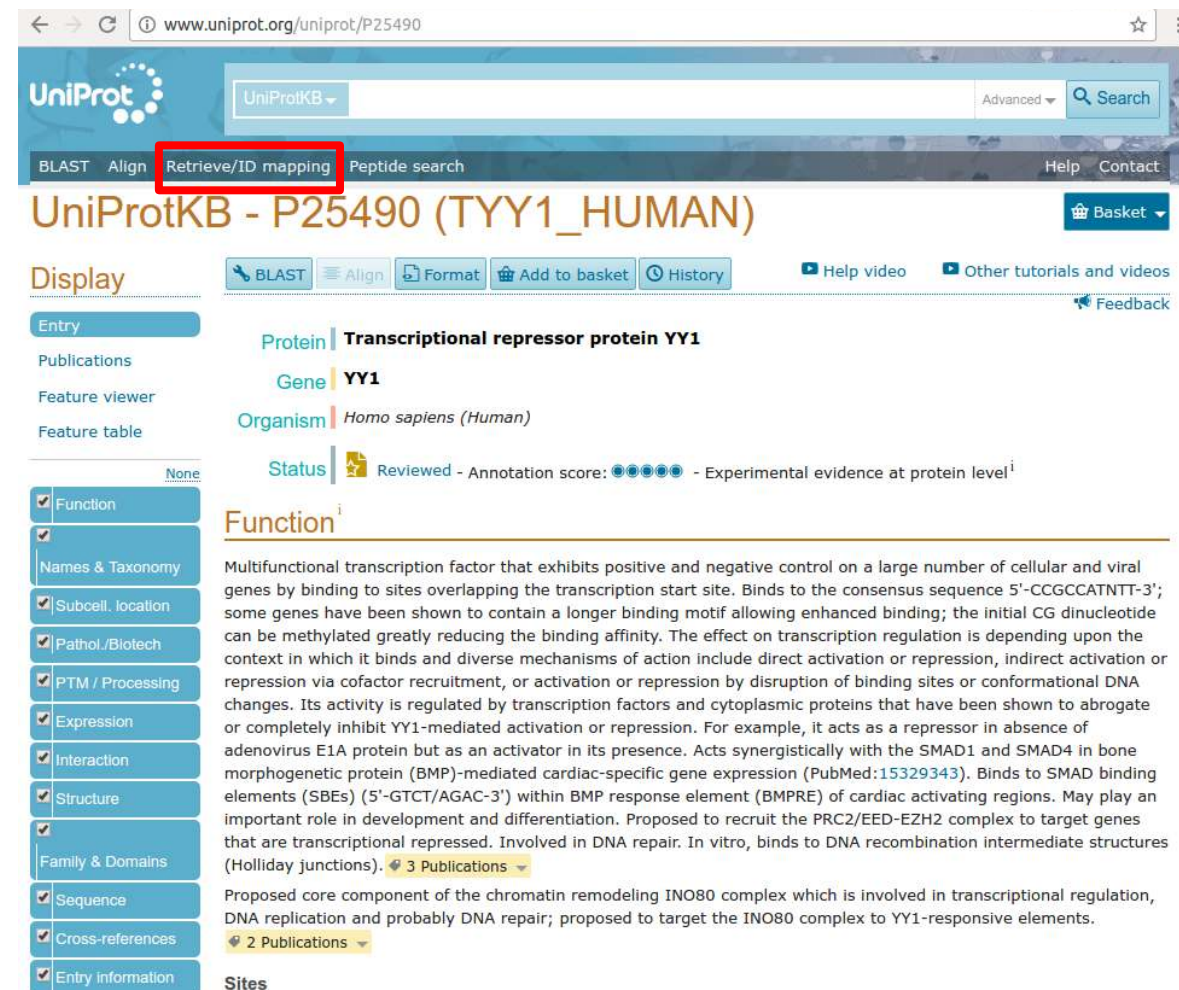
Базы данных, содержащие информацию по белкам



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



None

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

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).  3 Publications 

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

Sites

Лайфхак

NOYJV7_HUMAN

 Unreviewed (TrEMBL)

PDB - Protein Data Bank

- <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

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 EMDB 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
Clashescore		10
Ramachandran outliers		0.3%
Sidechain outliers		3.2%
RSRZ outliers		1.2%

Worse | Better

Percentile relative to all X-ray structures.

Percentile relative to X-ray structures of similar resolution.

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](#)

Базы данных, содержащие информацию по микроРНК



miRDB

miRTarBase

miRBase

- <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

```
5' guucc a uu au ucugu au
   cuuc acu aacauggga gcuu cuc c
   |||| ||| ||||| |||| |||
3' gaag uga uuguaccuu ugaag gag g
   ---u a uu cg ---u aa
```

Get sequence

1633 reads, 84.5 reads per million, 35 experiments

Deep sequencing

Confidence Feedback: Do you believe this miRNA is real?

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


Coordinates (GRCm38) chr3: [127545228-127545301](#) [+]

Overlapping transcripts antisense [ENSMUST00000029588](#); Larp7-201; intron 8

- <http://www.mirdb.org/miRDB/index.html>

Содержит информацию о последовательностях микроРНК и их, спрогнозированным с помощью биоинформатики, мишеням для геномов человека, мыши, крысы, собаки и курицы.

- Идентификаторы: название микроРНК. Например, hsa-miR-302b-3p.
- ✓ Приставка «mir» отделяется дефисом, вслед за ней следует номер, говорящий о порядке именования (mir-123 была открыта и названа раньше, чем mir-302).
- ✓ «mir-» обозначает пре-микроРНК, «MIR-» ген, кодирующий микроРНК, а «miR-» — для обозначения зрелой формы.
- ✓ К названию микроРНК с последовательностями, отличающимися на один или два нуклеотида, приписывается строчная буква (miR-123a и miR-123b).
- ✓ Пре-микроРНК, дающие начало на 100 % идентичным микроРНК, но локализованные в разных местах генома, имеют в названии цифру, отделенную дефисом (hsa-mir-194-1 и hsa-mir-194-2).
- ✓ Вид, из которого была выделена микроРНК, обозначается в названии трёхбуквенной приставкой (hsa-miR-123 человека).
- ✓ Когда две зрелые микроРНК образуются сразу с 3' и 5' концов исходной пре-микроРНК, к ним добавляется суффикс -3p или -5p. А если известен уровень экспрессии для этих микроРНК, тогда микроРНК с меньшей экспрессией 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	Seed Location	984, 1137, 1538, 1600, 3494
Target Score	100		
NCBI Gene ID	55432	GenBank Accession	NM_018566
Gene Symbol	YOD1	3' UTR Length	5171
Gene Description	YOD1 deubiquitinase		

3' UTR Sequence

```

1 CCTATGCATG AATGAGG6TT GAAGCCTACT ACCTCACACA TCCAGAAGGC TCTGG6TTTT
61 CCAATAAGCT ATGGTAACCC TAAAGAACAA AGGATACAAT GCTTGAACCA TCCTTTTAAC
121 TTAATAACAC TAAGACACTG AAATTCCTTG TTAAGATTAA AATTAGTGTG CAAGTTTACA
181 GATGTGTGTC TACAGTGGTA AACTGTACAT ACATGCCTCT TTCTGCTGGA GTGACAGAAT
241 AGGTGATCCT TGGACCTAC TGAAGCTGAC CTGAAGGTTG AGATTGAGTA TTATAAAGTA
    
```



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

- <http://mirtarbase.mbc.nctu.edu.tw/>
- Содержит информацию по экспериментально подтвержденным взаимодействиям микроРНК-мишень для 23 организмов.
- Идентификаторы: микроРНК человека hsa-miR-26b-5p имеет идентификатор MIRT029499.
- Содержит информацию по
 - ✓ Мишеням (23054 генов для 4076 микроРНК)
 - ✓ Последовательности
 - ✓ Ассоциированным заболеваниям
 - ✓ Ссылки на публикации и др.

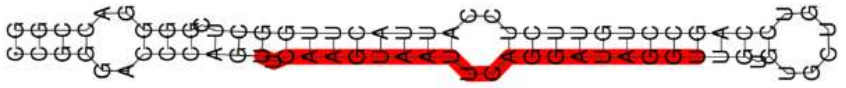
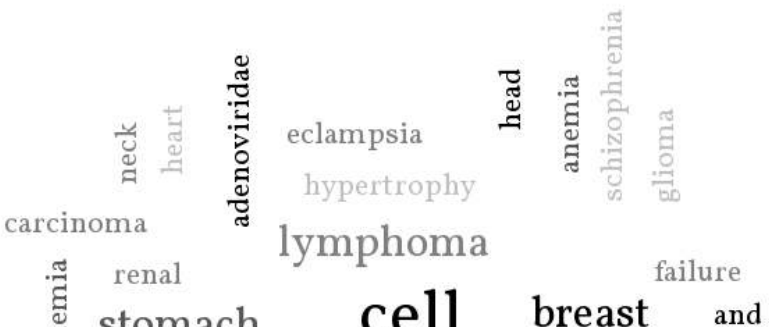
Информационная карточка микроРНК.

miRTarBase

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

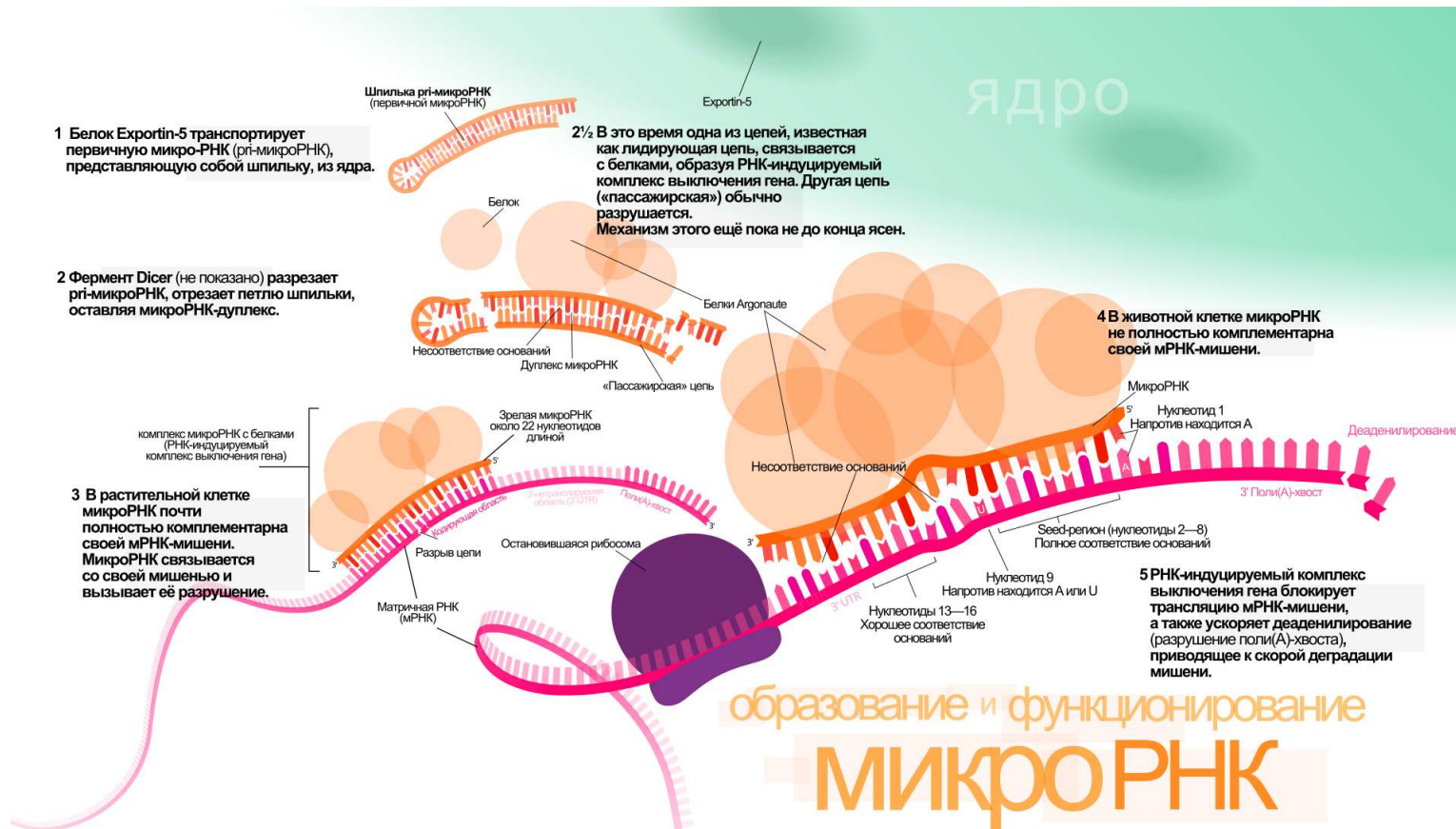
miRNA Target Gene Evidences Expression TCGA Gene Set Enrichments Network ERROR Report

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.

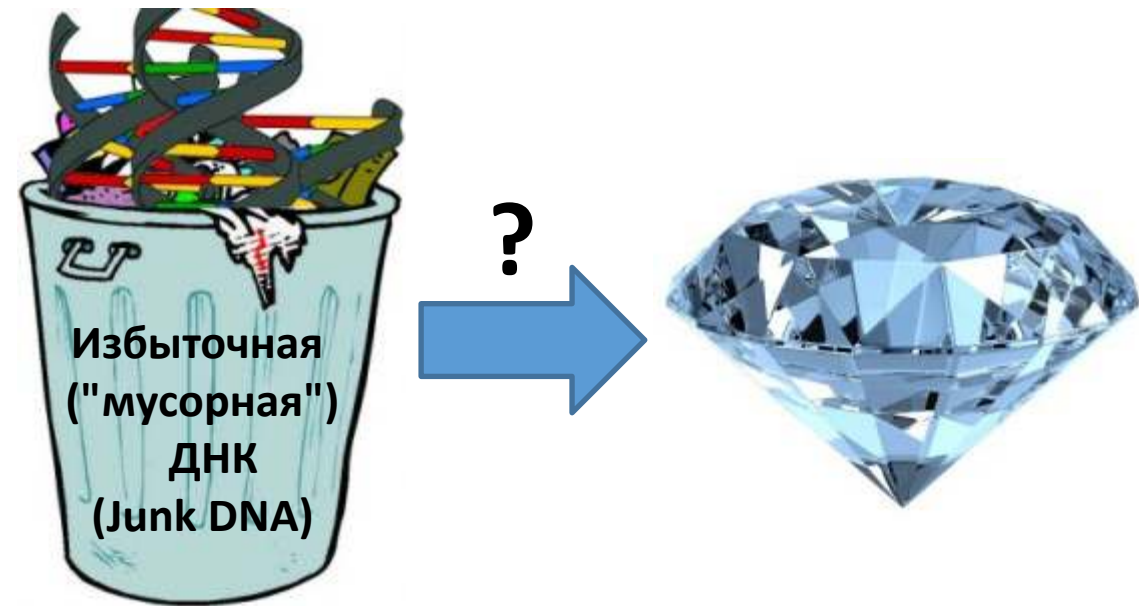
А какие еще классы некодирующих РНК Вы можете назвать?



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



Базы данных, содержащие информацию по метаболитам



ChEBI

PubChem



ChEBI

- Доступна через интернет по адресу <https://www.ebi.ac.uk/chebi/init.do>
- Посвящена малым химическим соединениям, как природного, так и искусственного происхождения.

Содержит информацию по

- ✓ функциям
- ✓ химической формуле
- ✓ молекулярному весу
- ✓ биологической роли
- ✓ онтологии (родство с другими химическими соединениями)

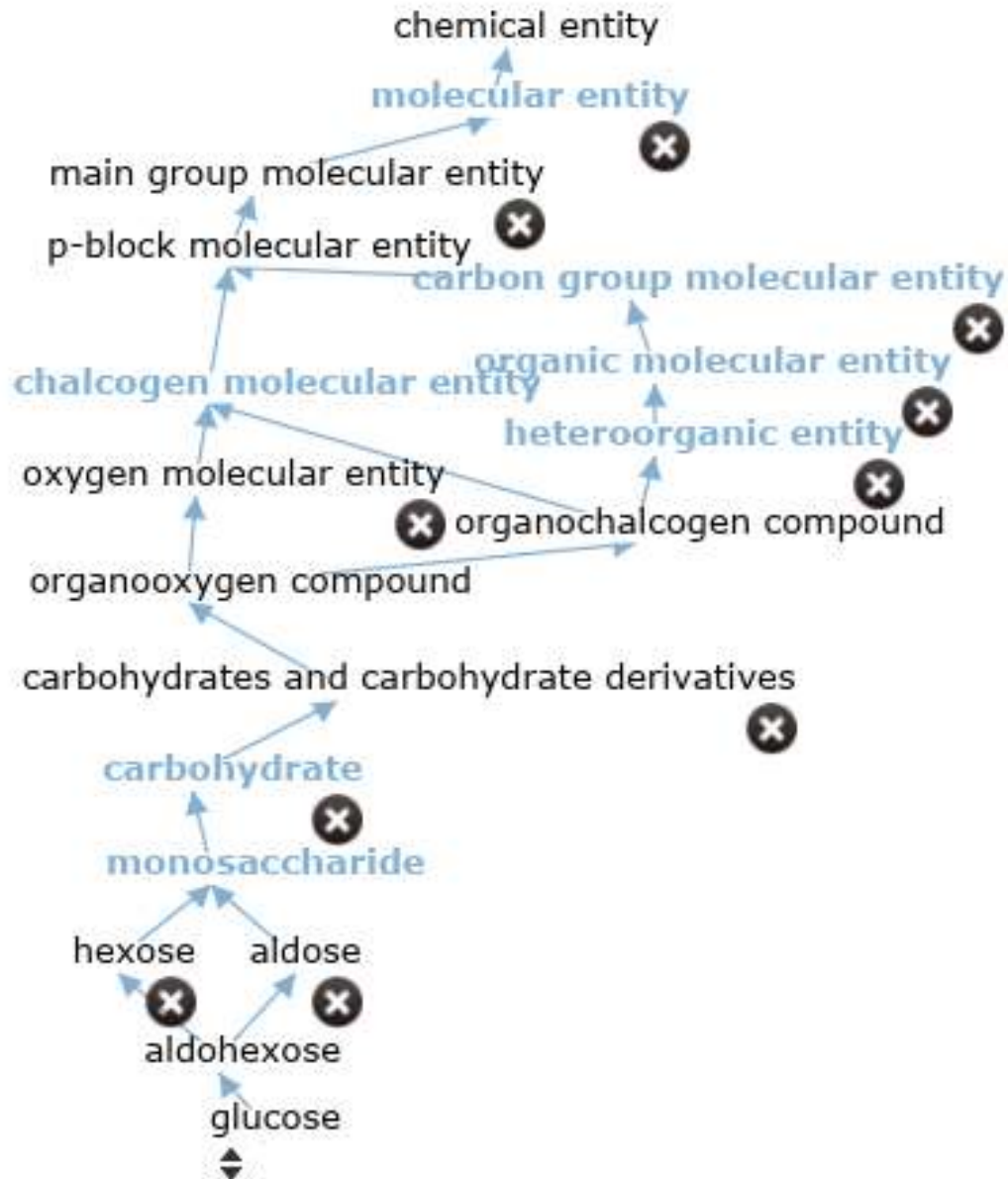
Информационная карточка метаболита

The screenshot shows the ChEBI website interface. At the top, there is a search bar with the text "CHEBI:153" and a "Search" button. Below the search bar, there are navigation links: Home, Advanced Search, Browse, Documentation, Download, Tools, About ChEBI, Preferences, and Submit. The main content area displays the title "CHEBI:15347 - acetone" and a tabbed interface with "Main" selected. The "Main" tab shows a chemical structure of acetone (H₃C-C(=O)-CH₃) and a table of properties:

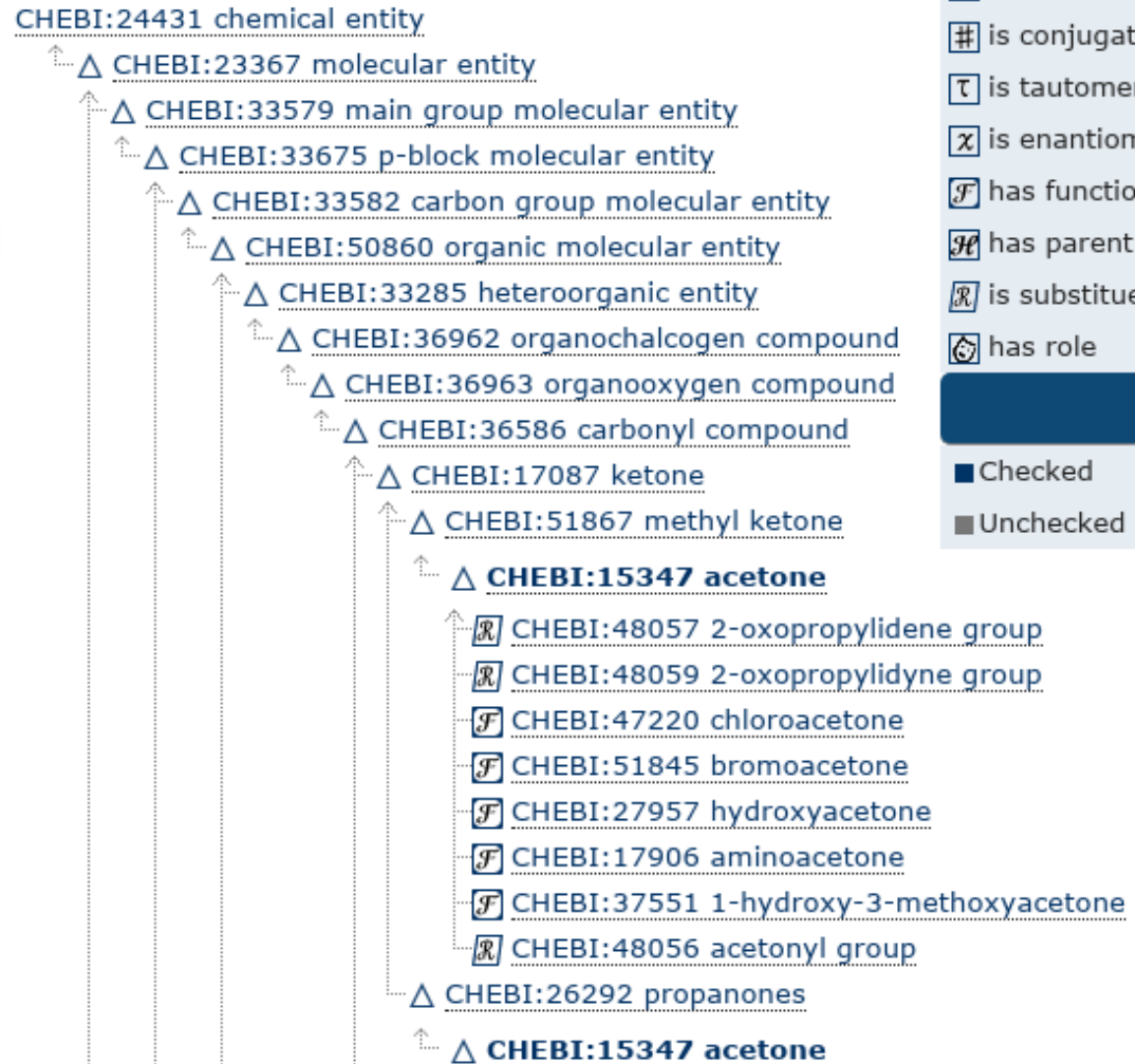
ChEBI Name	acetone
ChEBI ID	CHEBI:15347
Definition	A methyl ketone that consists of propane bearing an oxo group at C2.
Stars	☆☆☆ This entity has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs	CHEBI:40571, CHEBI:2398, CHEBI:13708, CHEBI:22182
Supplier Information	ZINC00895111, eMolecules:474422

Below the table, there are links for "Download Molfile" and "more structures >>". At the bottom, there is a "Wikipedia" section with a "License" button and a paragraph of text: "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

ChEBI



Онтология



Relationship Types

- Δ is a
- \diamond has part
- \square_b is conjugate base of
- $\square_{\#}$ is conjugate acid of
- \square_{τ} is tautomer of
- \square_{χ} is enantiomer of
- \mathcal{F} has functional parent
- \mathcal{H} has parent hydride
- \mathcal{R} is substituent group from
- \textcircled{C} has role

Status

- Checked
- Unchecked

PubChem

- Доступна через интернет по адресу <https://pubchem.ncbi.nlm.nih.gov/compound/>
- Посвящена малым химическим соединениям, как природного, так и искусственного происхождения.
- Содержит информацию по
 - ✓ синонимам
 - ✓ функциям и физико-химическим свойствам
 - ✓ химической формуле
 - ✓ молекулярному весу
 - ✓ биологической роли
 - ✓ производителям и коммерческим названиям
 - ✓ патентам

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

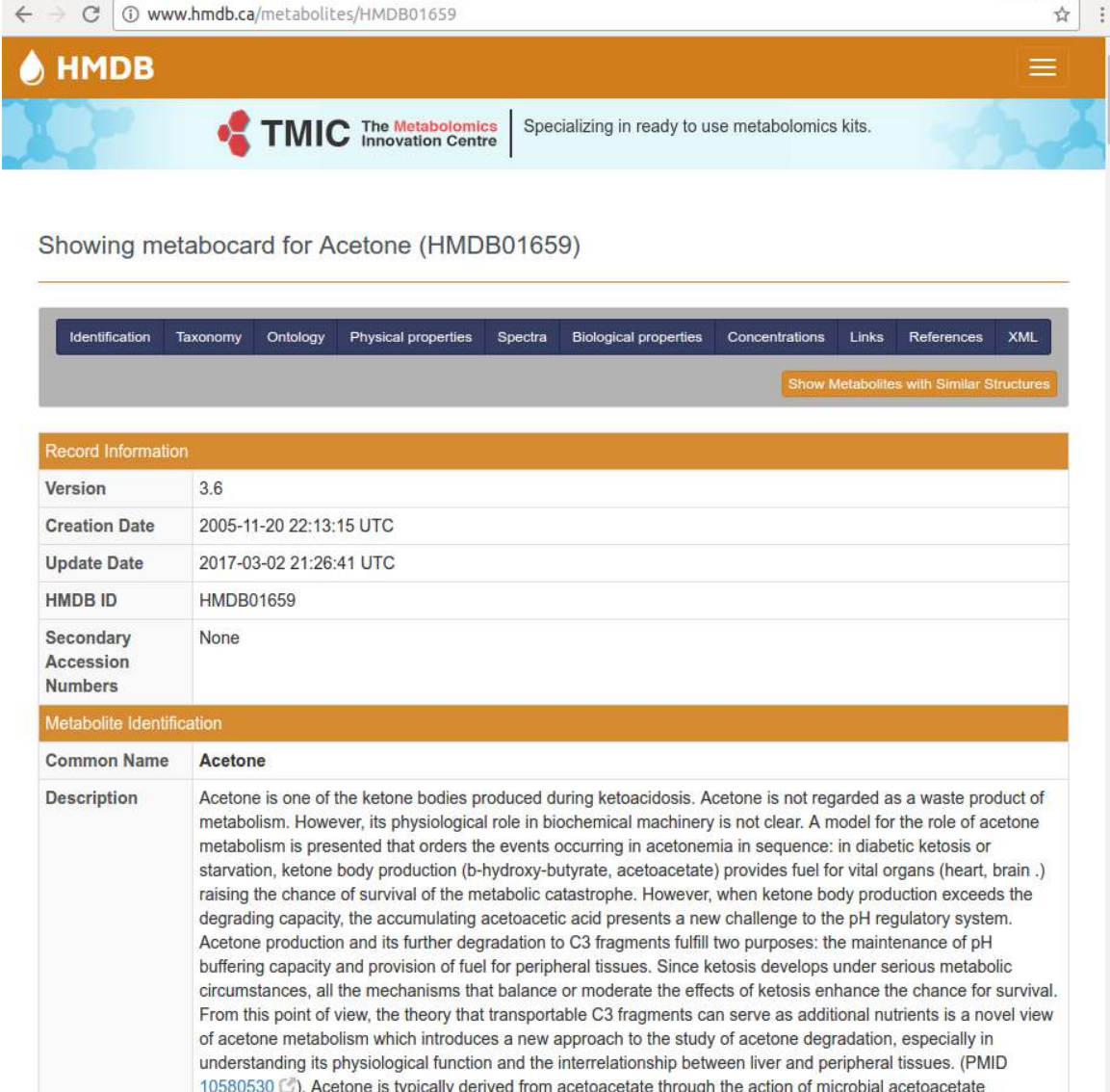
HMDB

- Доступна через интернет по адресу <http://www.hmdb.ca/>

Посвящена метаболитам человека.

- 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 C). Acetone is typically derived from acetoacetate through the action of microbial acetoacetate

Базы данных, содержащие информацию по заболеваниям.

OMIM[®]

ICD-11

 DISEASE
ONTOLOGY

OMIM

- <https://www.omim.org/>
- Наследуемые заболевания человека
- Содержит информацию по
 - ✓ синонимам
 - ✓ ассоциациям с генами
 - ✓ наследуемости
 - ✓ биохимических особенностях
 - ✓ патогенезе
 - ✓ животным моделям и др.

Идентификаторы: #1258530 (или MIM:1258530).

* - ген

- фенотип, не один локус

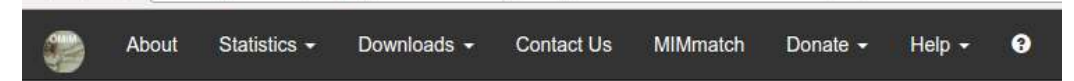
+ - ген/фенотип (напр. + 159555)

%- фенотип, неизвестны гены (% 189800)

^ - устаревший идентификатор

1,2,6 – аутосомы,
3 – X-хромосома,
4 – Y-хромосома,
5 - митохондрии

Информационная карточка заболевания



diabetes



Options

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

External Links

Protein

Clinical Resources

Clinical Trials
Genetic Alliance
Genetics Home Reference
GTR

Animal Models

Cell Lines

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

ICD11 - International Statistical Classification of Diseases and Related Health Problems 11th Revision

- <https://icd.who.int/>
- Содержит классификацию заболеваний человека, принятую международным сообществом и используемую практикующими врачами.

Идентификаторы: от 1A00.00 до ZZ9Z.ZZ

Номер основной категории, от 1 до Z (кроме символов O, I)

Подробнее в Reference guide раздел 2.2.1

- Нелирическое отступление:
- Как найти идентификаторы в текстовых редакторах с помощью ctrl+F?

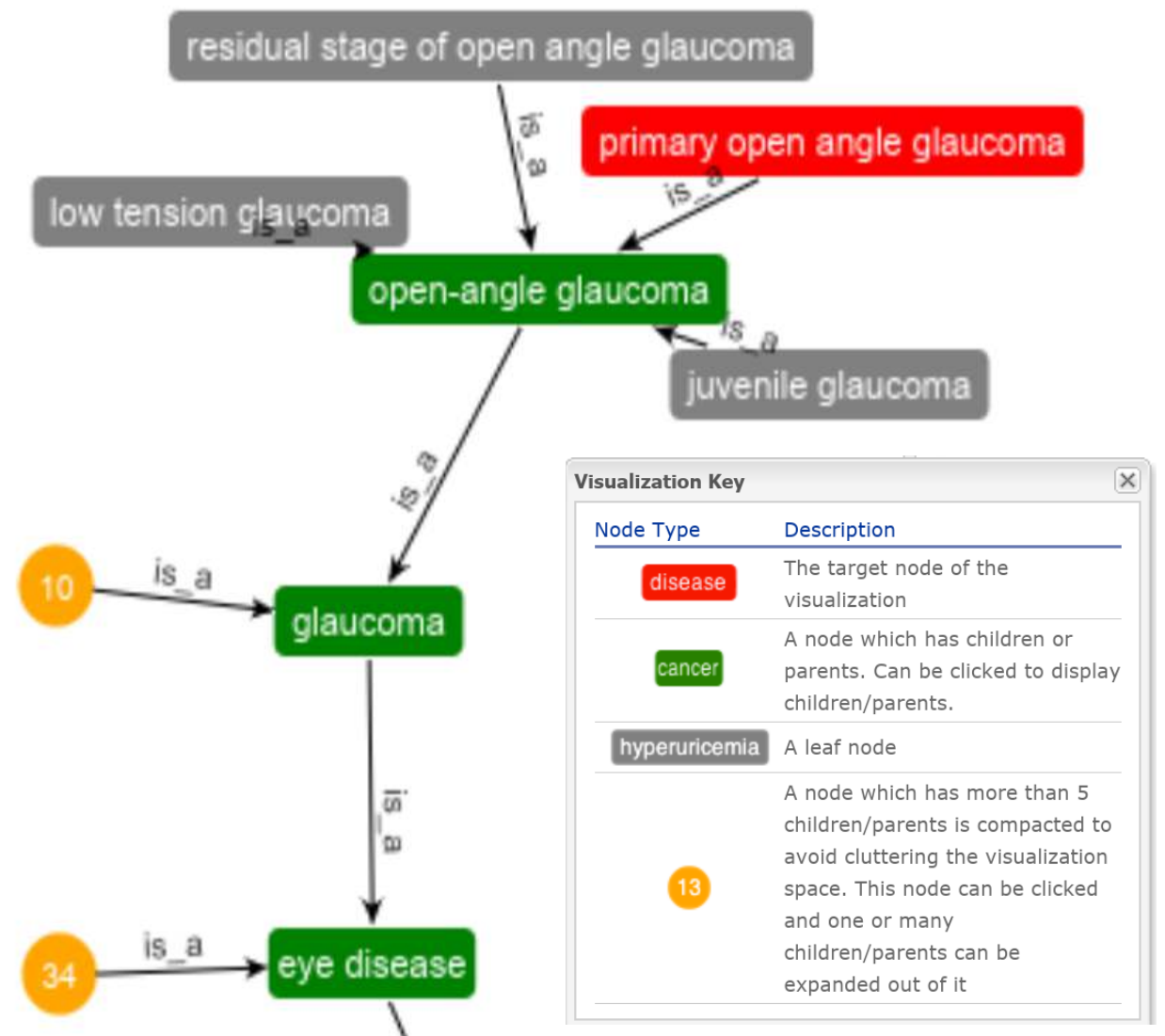
Использовать регулярные выражения (regex) или подстановочные знаки. Пример : `\w[A-Z]\d\w\.\w{2}\w{4}\.\w{2}`

The screenshot shows the ICD-11 for Mortality and Morbidity Statistics website. The search bar contains 'Diabetes mellitus'. The results are displayed in a tree view under '05 Endocrine, nutritional or metabolic diseases'. The 'Diabetes mellitus' category is highlighted, showing sub-categories like 'Type 1 diabetes mellitus', 'Type 2 diabetes mellitus', and 'Diabetes mellitus in pregnancy'. On the right side, there is a detailed view for 'Diabetes mellitus', including its parent category 'Endocrine diseases', a description 'A metabolic disorder with hetero both.', and a list of 'Coded Elsewhere' items such as 'Diabetes mellitus in pregnancy' and 'Neonatal diabetes mellitus'.

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.

Онтология



Оцените, сколько в настоящее время известно заболеваний?



Оцените, сколько в настоящее время известно заболеваний?
В МКБ-10 описано около 12,5 тысяч заболеваний (МКБ-10
одобрена ВАЗ в 1990 году), .
В МКБ-11 - около 16 тысяч (одобрена в мае 2019).



Базы данных, содержащие информацию по биологическим путям/процессам.



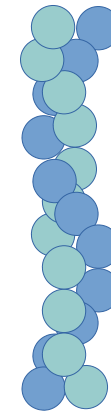
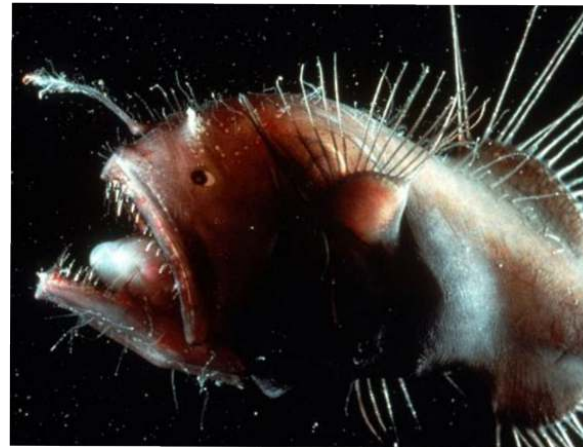
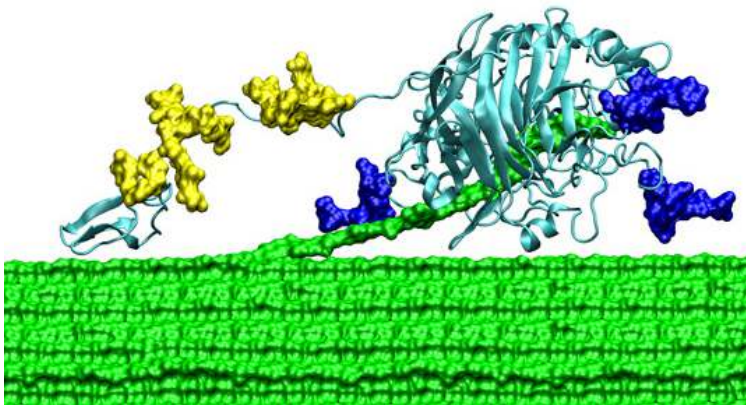
Что такое биологический процесс?

функция

Биохимическая
активность

Биологическая цель

Структура



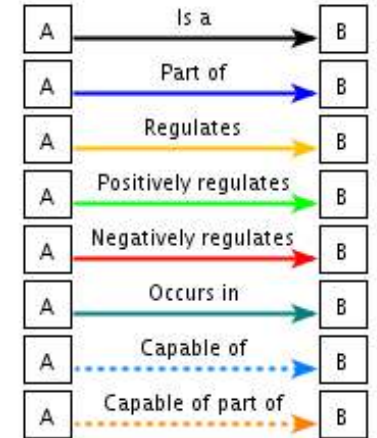
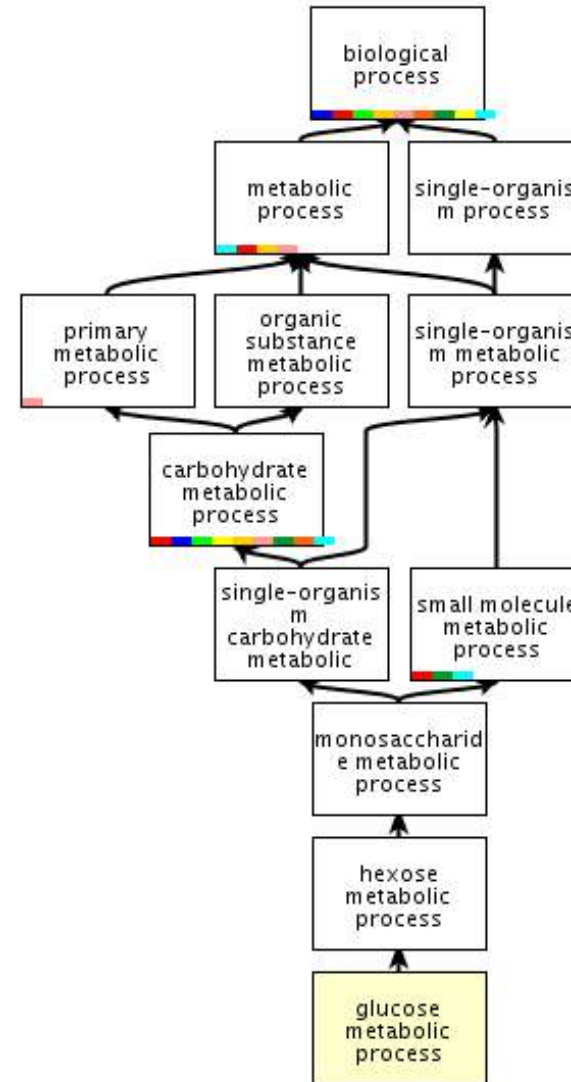
Molecular function
(MF)

Biological process
(BP)

Cellular component
(CC)

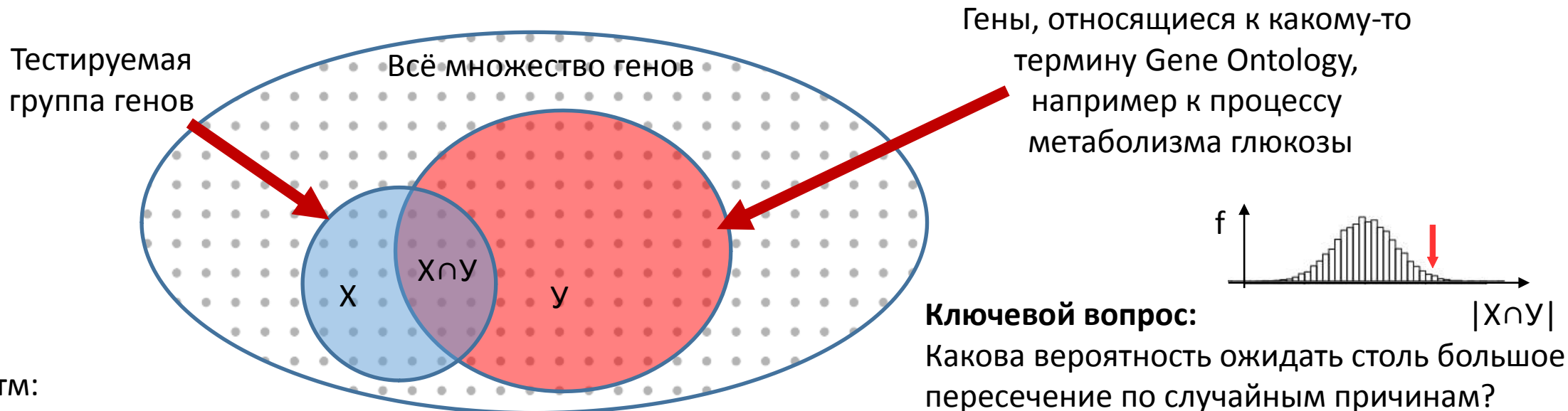
Gene Ontology

- <http://www.geneontology.org/>
- Является онтологией, описывающей биологические процессы (BP), клеточные компоненты (CC) и молекулярные функции (MF). Имеется привязка к генам различных организмов).
- Идентификаторы: GO:семизначное число. Например, glucose metabolic process имеет идентификатор GO:0006006.
- Содержит:
 - ✓ синонимы
 - ✓ определение
 - ✓ ассоциации с генами (<https://www.ebi.ac.uk/GOA>)



Поиск сверхпредставленных биологических процессов

Сверхпредставленность процесса (фенотипа, заболевания) – повышенная представленность генов, приписанных данному процессу в исследуемой группе генов. Названия на английском: Over-represented biological processes, Gene Enrichment Analysis, Gene Ontology Enrichment Analysis.



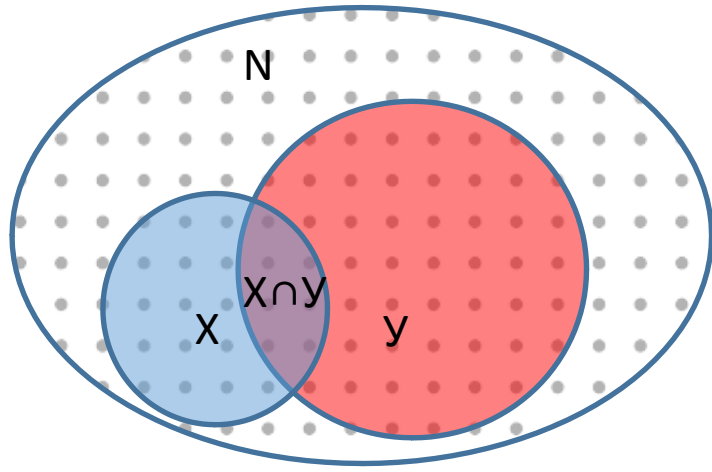
Алгоритм:

1. Выявляется тестируемая группа генов X (например, с помощью эксперимента типа случай/контроль).
2. Выявляется набор генов Y , относящихся к какому-то термину Gene Ontology.
3. С помощью точного критерия Фишера (гипергеометрического распределения) или других статистических методов оценивается вероятность обнаружить такое же или большее пересечение между множествами по случайным причинам (p -value). При этом учитываются такие величины как $X \cap Y$, X , Y и число всех генов.
4. Процедура из пунктов 2, 3 повторяется для всех Gene Ontology терминов. Так как всего Gene Ontology терминов много (порядка 10 тысяч), необходимо делать поправку на множественное сравнение, например, Бенджамини-Хокберга.

Точный критерий Фишера

Что наблюдается? Таблица сопряжённости (2x2)

	Число генов в наборе X	Число генов не в наборе X	Полное число генов
Число генов в наборе Y	$ X \cap Y = 5$	$ Y - X \cap Y = 1$	$ Y = 6$
Число генов не в наборе Y	$ X - X \cap Y = 1$	$N - X \cup Y = 23500 - 6 - 6 + 5 = 23493$	$N - Y = 23500 - 6 = 23494$
Полное число генов	$ X = 6$	$N - X = 23500 - 6 = 23494$	$N = 23500$



Что ожидается по случайным причинам?

$$N(N-1)(N-2)\dots(N-|X|+1) = \frac{N!}{(N-|X|)!}$$

Ген А, ген Б, ... ген В

...

Ген В, ген А, ... ген В



$$|X|!$$

$$\frac{N!}{(N-|X|)! X!} = C_N^{|X|} \text{ Полное число вариантов } X_{\text{случайный}}$$

Сколько вариантов генов из Y_{случайный}

окажутся в X_{случайный}, если известно,

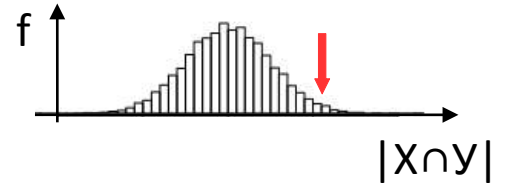
что их $|X \cap Y|$ (как в наблюдении)?

$$C_{|Y|}^{|X \cap Y|}$$

Сколькими способами можно заполнить оставшиеся места в X_{случайный} генами не из Y_{случайный} $C_{N-|Y|}^{|X|-|X \cap Y|}$

Какова ожидаемая частота (вероятность)?

$$\frac{C_{|Y|}^{|X \cap Y|} \cdot C_{N-|Y|}^{|X|-|X \cap Y|}}{C_N^{|X|}}$$



$$p\text{-значение} = \frac{C_6^5 \cdot C_{23494}^1}{C_{23500}^6} + \frac{C_6^6 \cdot C_{23494}^0}{C_{23500}^6} = 6,03 \cdot 10^{-19}$$

Как посчитать в R?

```
x<-matrix(c(5, 1, 1, 23493), 2, 2)
fisher.test(x, alternative = "greater")
```

Пакеты Bioconductor: GOStats, TopGO

KEGG - Kyoto Encyclopedia of Genes and Genomes

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

Информационная карточка биологического процесса.

KEGG PATHWAY: hsa00010 Help

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

Reactome

- <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.



Glucose metabolism

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

Locations in the PathwayBrowser

The screenshot shows the Reactome PathwayBrowser interface. On the left is a sidebar with an "Event Hierarchy" tree, where "Metabolism" and "Metabolism of carbohydrates" are expanded. The main area displays a complex metabolic pathway diagram with various molecules and reactions. Several sub-pathways are highlighted with colored boxes: "Digestion of dietary carbohydrate" (blue), "Galactose catabolism" (red), "Fructose metabolism" (orange), "Glucose metabolism" (green), "Penicillamine metabolism" (purple), "5-Phosphoribose 4-epimerase (Biosynthesis)" (yellow), and "Catabolism of glucuronate to xylulose-5-phosphate" (brown). At the bottom, there is a "Description" tab, a "Molecules" list, and a "Summation" section. The "Summation" section contains a text box with a summary of the pathway's function.

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

inner mitochondrial mem

Какой процесс отвечает за «потерю хвостика» головастиком?



Какой процесс отвечает за «потерю хвостика» головастиком?

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



План лекции

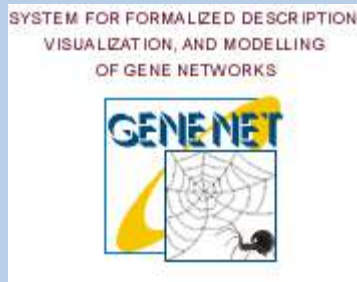
1. Введение.
2. Обзор Интернет-ресурсов, содержащих разнообразную информацию о генных сетях.
3. Ресурсы, интегрирующие биологическую информацию из разнородных источников и представляющие ее в виде генных сетей: ANDSystem, STRING, GeneMania, Pathway Commons.
4. Практическое применение инструментов интеграции.

Два типа инструментов интеграции

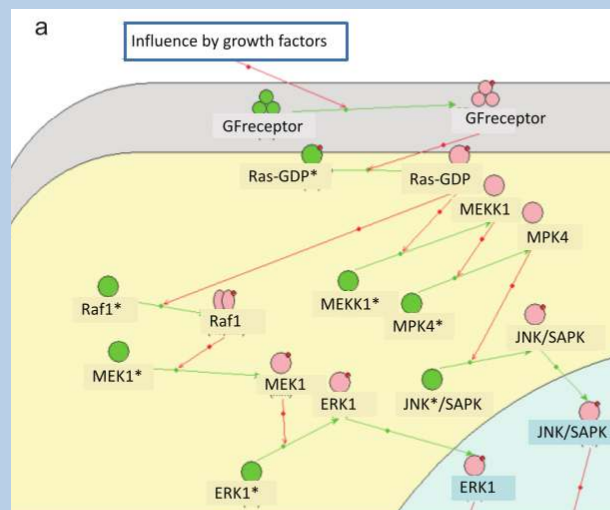
1. **Ручная реконструкция генных сетей** экспертами с использованием специальных программных средств – редакторов генных сетей.



WIKIPATHWAYS
Pathways for the People



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



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

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

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

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

Система ANDSystem включает

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

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

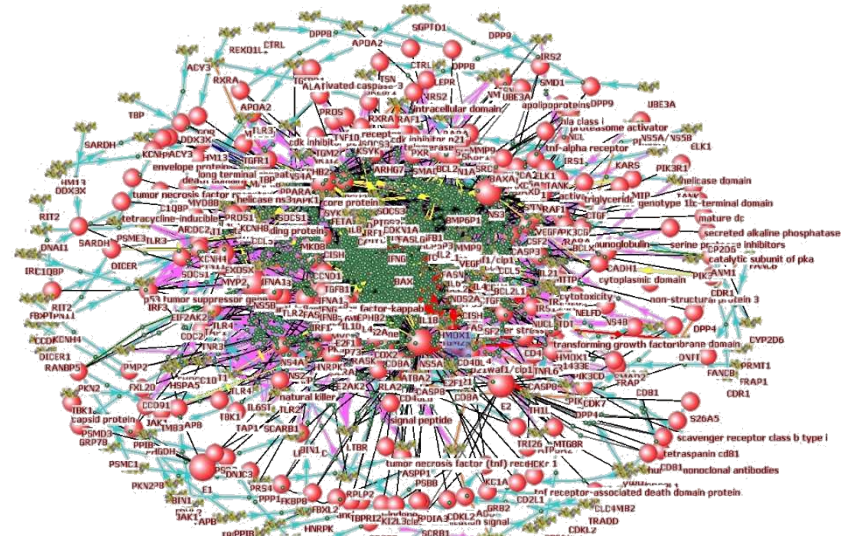
- молекулярно-генетическими системами и процессами
- заболеваниями и фенотипическими признаками
- факторами окружающей среды и др.



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

● белки
→ усиление активности
→ подавление активности
→ образование комплекса

The network diagram shows interactions between several proteins: NSSA, NSSB, TRAF2, JNK, TNF-alpha, IKKalpha, and NK-kB. NSSA is shown to interact with TRAF2 and JNK. NSSB is shown to interact with TRAF2, JNK, TNF-alpha, and IKKalpha. TNF-alpha is shown to interact with JNK and IKKalpha. NK-kB is shown to interact with TNF-alpha. The interactions are color-coded: green for activation, yellow for inhibition, and grey for complex formation.



Pathway Commons



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

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

Pathway Commons интегрирует информацию из 24 баз данных

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

Текстовое поле для ввода названия интересующего гена

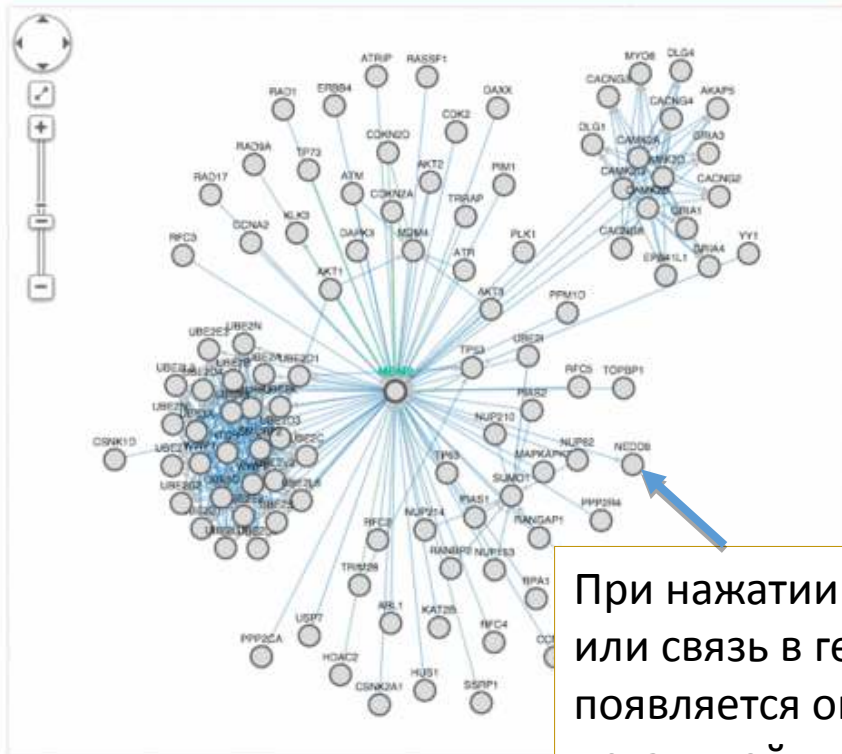
Добавление в сеть данных по изменениям, наблюдающимся при раке

Настройки для фильтрации сети по связям и вершинам

PCViz Pathway Commons Network Visualizer

Genes of interest

MDM2 +



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

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

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, RIS2

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

Details Settings Context

Interaction types

- 362 controls state change
- 409 controls expression

Number of genes (468)

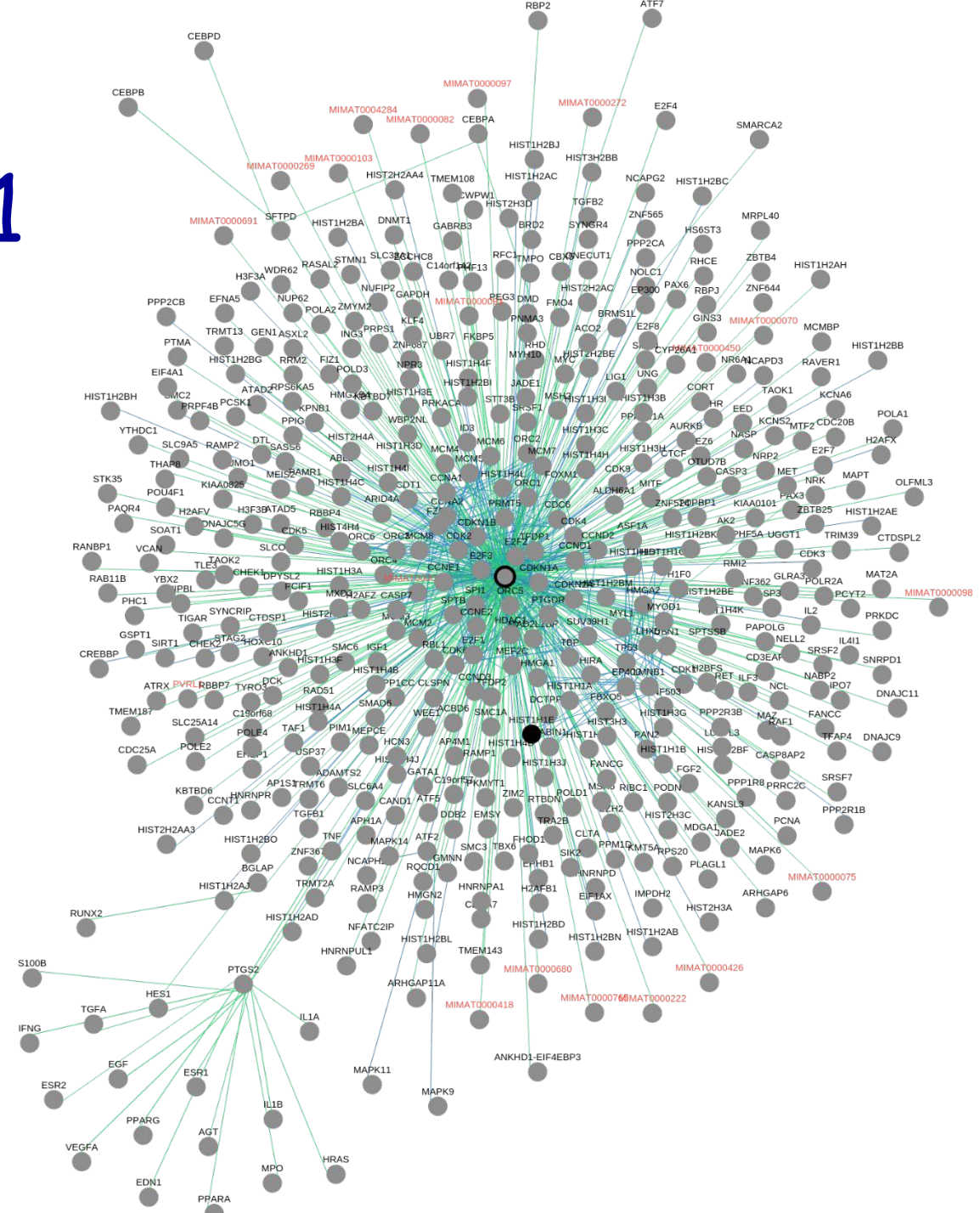
Slide left to decrease the number of genes

Query type: Neighborhood

Пример генной сети для гена ретинобластомы RB1

Пример загрузки генной сети в формате SIF

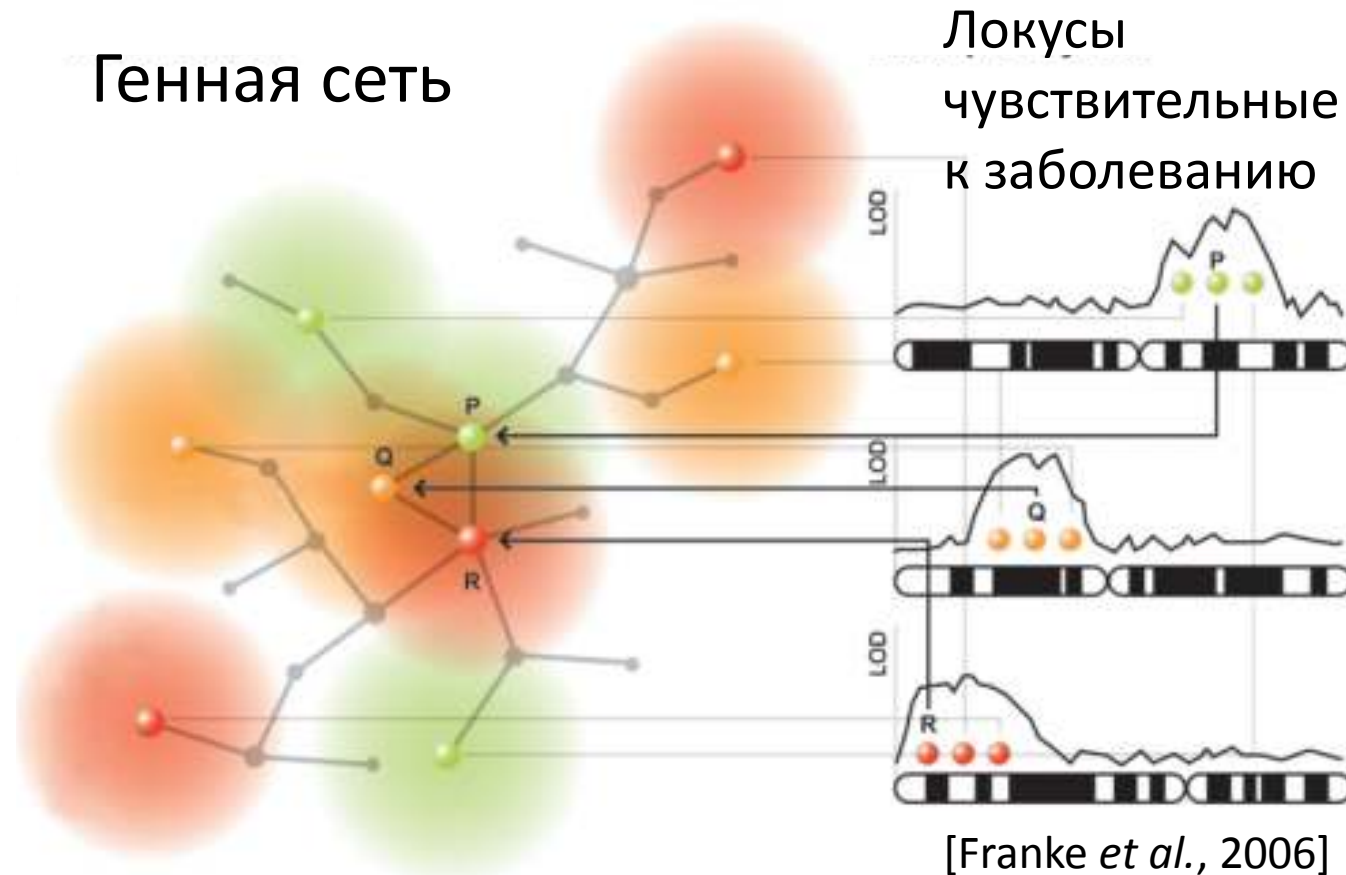
AATF	in-complex-with	E2F1
AATF	in-complex-with	E2F2
AATF	in-complex-with	E2F3
AATF	in-complex-with	RB1
AATF	interacts-with	RB1
AATF	in-complex-with	TFDP1
ABCD3	interacts-with	DYRK1A
ABCD3	interacts-with	DYRK1B
ABCD3	interacts-with	RB1
ABHD10	interacts-with	FOXK1



План лекции

1. Введение.
2. Обзор Интернет-ресурсов, содержащих разнообразную информацию о генных сетях.
3. Ресурсы, интегрирующие биологическую информацию из разнородных источников и представляющие ее в виде генных сетей: ANDSystem, STRING, GeneMania, Pathway Commons.
4. Практическое применение инструментов интеграции.

Поиск исходных вершин для реконструкции генных сетей



<https://www.sciencedirect.com/science/article/pii/S0002929707639226>

1. Избавление от фактора сцепленности генов в хромосомах при использовании информации об ассоциациях аллелей с заболеванием
2. Нахождение генов, мутации в которых являются первопричинами заболеваний

OPEN

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

Juthatip Manisom, Supaporn Khamchun, Arada Vinaiphat & VisithThongboonkerd

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 performed global protein network analysis of those altered proteins. Subsequently, overexpression of a protein, which was one of the central nodes of such protein-protein interactions network, was performed. Moreover, functional investigations were performed to address functional significance of the central-node protein and its associated partners in kidney stone disease.

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

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 vitro* models [8, 9] have provided some insights in this

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

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

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

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.



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*}, Nikolaus Schultz², Chris Sander^{1*}

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

* Current address: Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland * serbaba@mskcc.org (YS); chs@сандerlab.org (CS)



CrossMark

click for updates

OPEN ACCESS

Citation: Şenbabaoğlu Y, Sümer SO, Sánchez-Vega F, Bemis D, Criello G, Schultz N, et al. (2016) A Multi-Method Approach for Proteomic Network Inference in 11 Human Cancers. *PLoS Comput Biol* 12(2): e1004765. doi:10.1371/journal.pcbi.1004765

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

Received: June 7, 2015

Accepted: January 20, 2016

Published: February 29, 2016

Copyright: © 2016 Şenbabaoğlu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The pan-cancer RPPA dataset is available from The Cancer Proteome Atlas at <http://pan.cancerproteome.org/>. Design: <https://doi.org/10.1371/journal.pcbi.1004765>.

Funding: CS and YS were supported by the National Institutes of General Medical Sciences (P15GM103504) and the National Cancer Institute (U24CA143840). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Abstract

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

Author Summary

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

Анализ протеомных экспериментов, направленных на выявление белков в протеоме мочи чувствительных к различному содержанию NaCl в пище для человека

Период эксперимента	1- 35 сутки, или 1-5 неделя	36-70 сутки, или 6-9 неделя	71-75 сутки, или 10 неделя	76-98 сутки, или 11-14 неделя
Уровень потребления соли (г/сутки)	12	9	12	6

Примечание: норма потребления соли для человека находится в пределах 4-12 г/сутки.

Запись результатов протеомных экспериментов для проведения кластерного анализа

Недели эксперимента и уровень потребления соли														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
12 г/сут					9 г/сут					12 г/сут	6 г/сут			

Динамика появления белков в моче

Белок А

1	0	0	0	0	0	1	1	1	1	1	0	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---

Белок В

1	0	1	0	0	0	1	0	1	1	1	0	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---

Запись результатов протеомных экспериментов для проведения кластерного анализа

Недели эксперимента и уровень потребления соли														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
12 г/сут					9 г/сут					12 г/сут	6 г/сут			

Динамика появления белков в моче

Белок А

0,83	0,16	0	0	0,33	0,16	1	0,83	1	1	0,83	0	0,83	1
------	------	---	---	------	------	---	------	---	---	------	---	------	---

Белок В

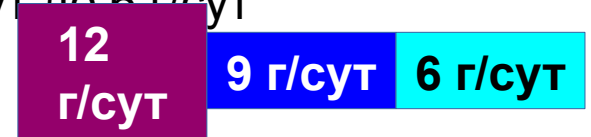
1	0	0,83	0	0	0	1	0,16	1	1	0,66	0,16	1	1
---	---	------	---	---	---	---	------	---	---	------	------	---	---

Манхэттенское расстояние между векторами: $D = \sum |A_i - B_i|$

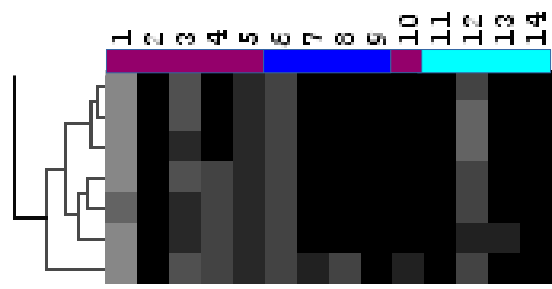
Доля испытуемых →
у которых присутствует
белок



← Номера недель, уменьшение
концентрации натрия в пище с
12 г/сут до 6 г/сут

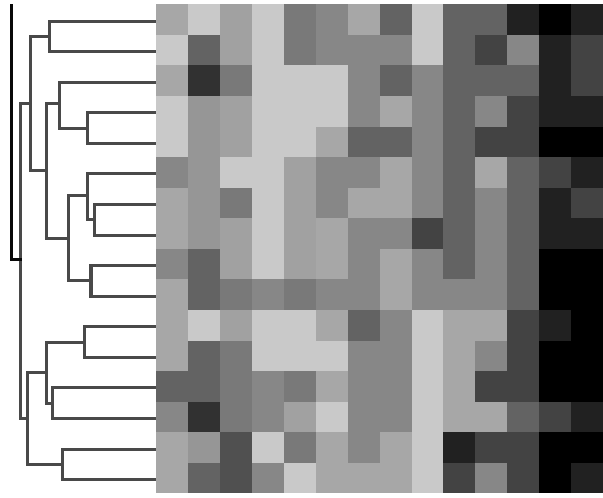


кластер 9



UBP15_HUMAN
LY75_HUMAN
HTFB2_HUMAN
AKA11_HUMAN
MYO5C_HUMAN
MYO5A_HUMAN
MYO6_HUMAN

кластер 83



CSPG4_HUMAN
MEN1_HUMAN
CO1A1_HUMAN
CD14_HUMAN
MGA_HUMAN
GELS_HUMAN
TETN_HUMAN
CYTM_HUMAN
CATD_HUMAN
SDCB1_HUMAN
CEL_HUMAN
CSF1_HUMAN
CADM1_HUMAN
LAIR1_HUMAN
DPP4_HUMAN
ACTB_HUMAN

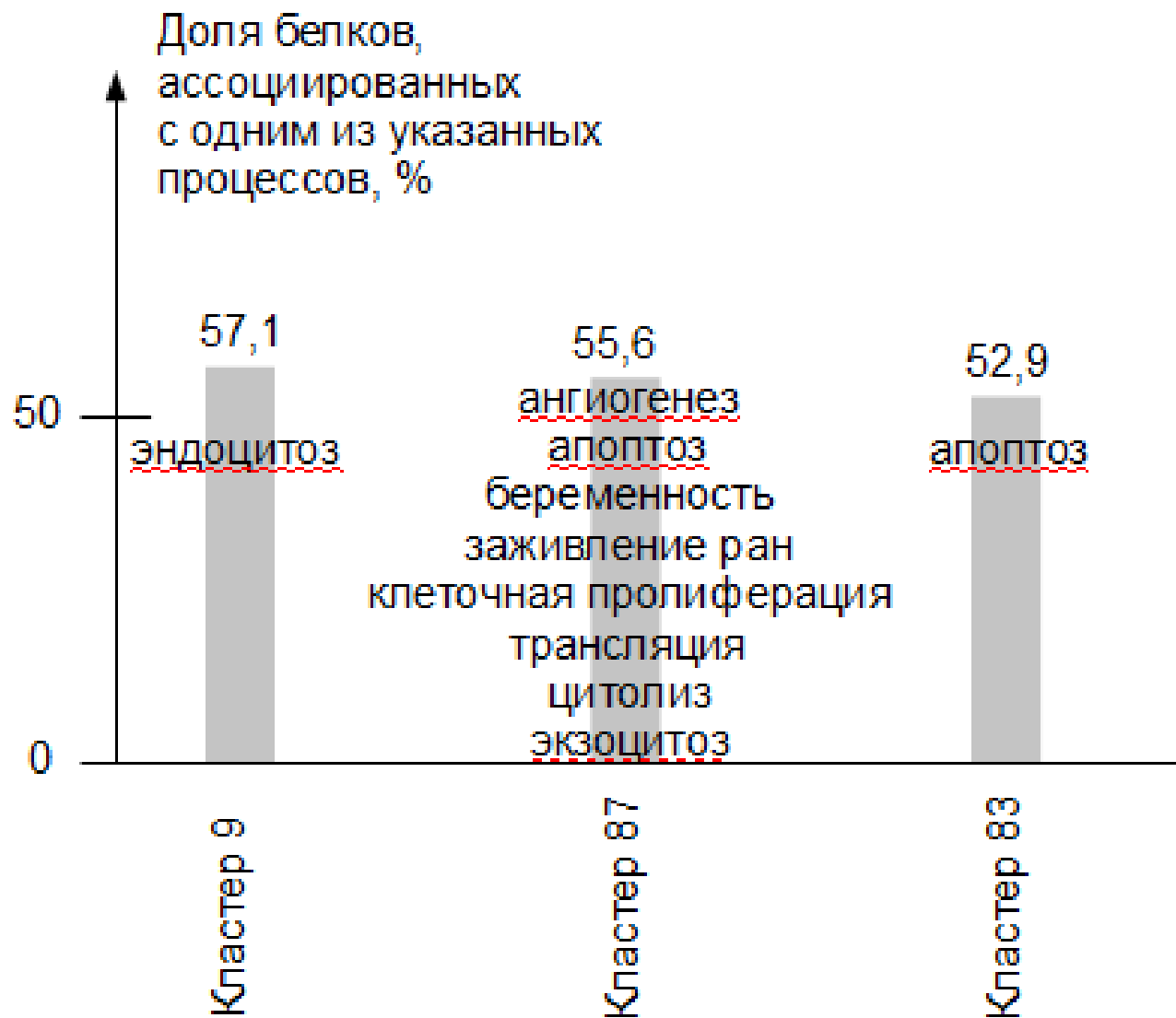
кластер 87



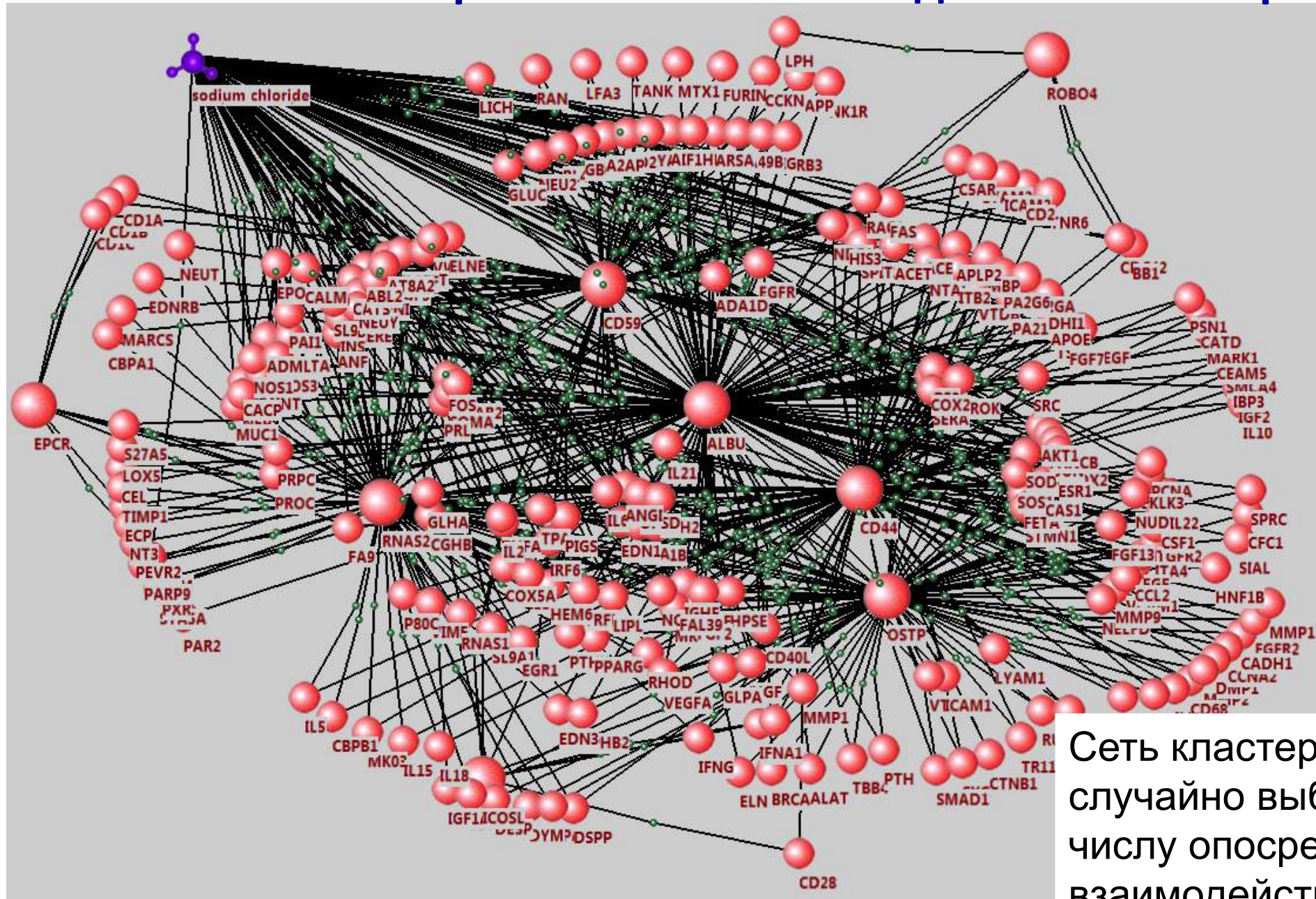
CD59_HUMAN
EPCR_HUMAN
ICOSL_HUMAN
RHAS2_HUMAN
CD44_HUMAN
YIPF3_HUMAN
ALBU_HUMAN
OSTP_HUMAN
ROBO4_HUMAN

Иерархическое
дерево было разбито
на кластеры
включающие не
более 30 элементов

Поиск кластеров ассоциированных с процессами



Построение сетей по найденным кластерам

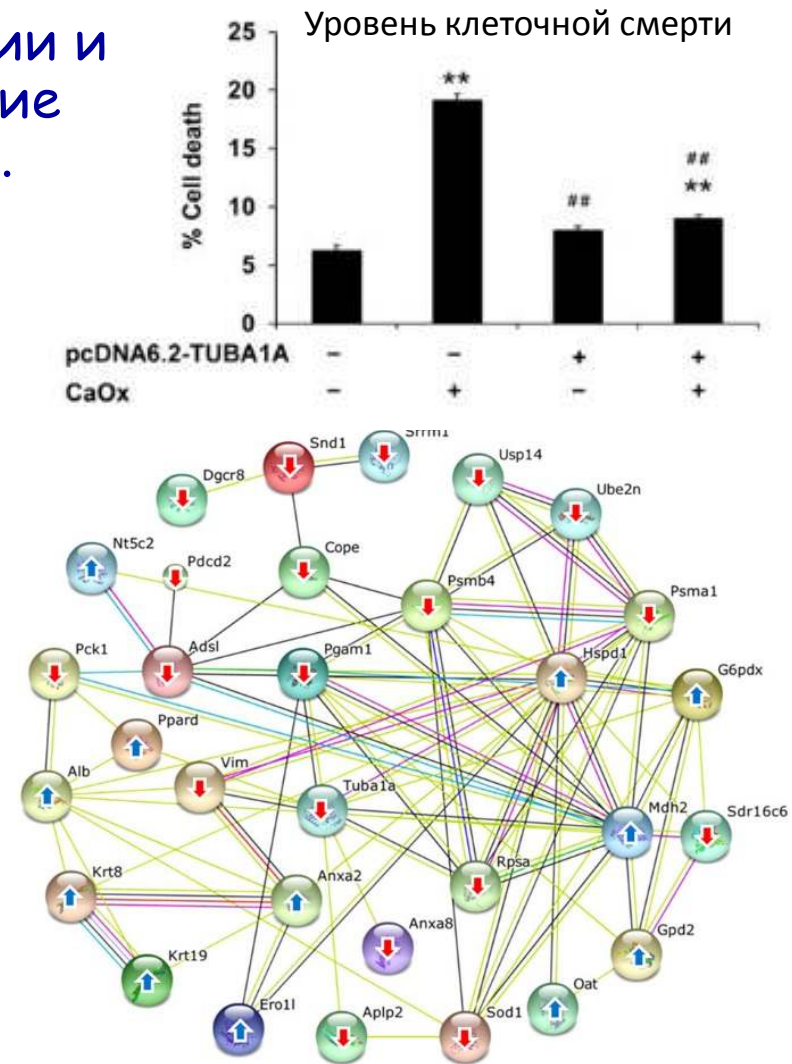
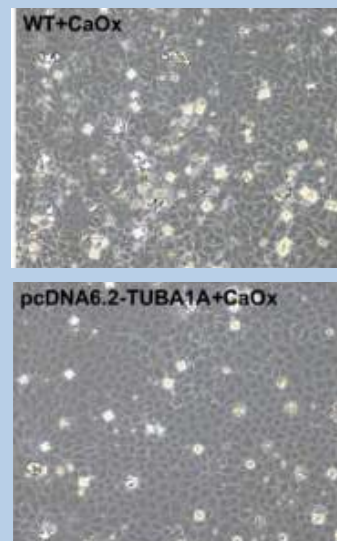


Сеть кластера 87 отличается от случайно выбранных сетей по числу опосредованных взаимодействий ($p=0.0175$)

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.

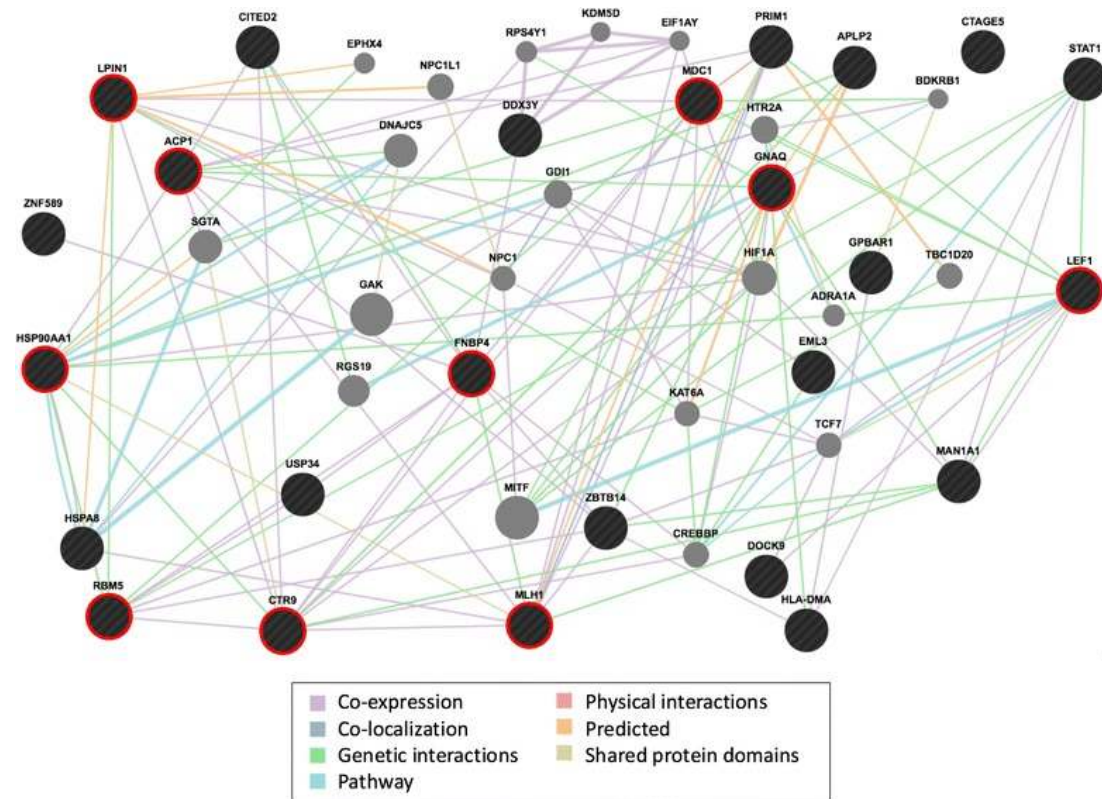
Фазово-контрастная микроскопия почечных трубчатых клеток



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.

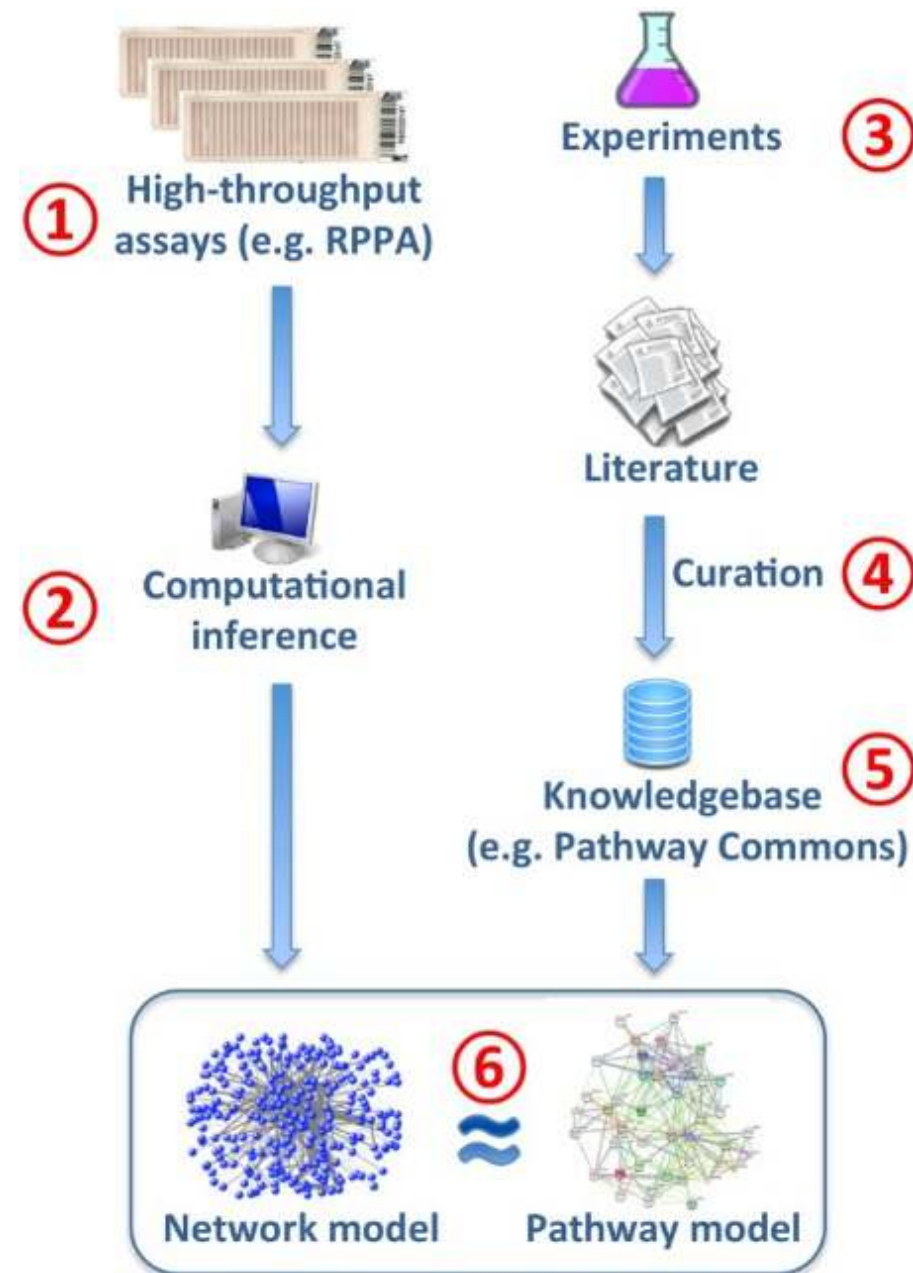
Воздействие загрязненного воздуха связано с неблагоприятными последствиями для здоровья, включая увеличение сердечно-легочной заболеваемости и смертности, а также увеличение риска рака легких. Был проведен полно-геномный анализ уровней экспрессии РНК крови 63 некурящих сотрудников парковочных терминалов 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 генной сети (реконструированной с использованием GeneMANIA) позволил выявить 25 взаимосвязанных генов, которые были общими для всех трех загрязнений. Многие из них ассоциированы с раком, а также вовлечены в связывание металлов и регуляцию апоптоза. Эти результаты описывают молекулярную взаимосвязь между воздействием загрязненного воздуха и неблагоприятными последствиями для здоровья.



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

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

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



План лекции

1. Введение.
2. Обзор Интернет-ресурсов, содержащих разнообразную информацию о генных сетях.
3. Ресурсы, интегрирующие биологическую информацию из разнородных источников и представляющие ее в виде генных сетей: ANDSystem, STRING, GeneMania, Pathway Commons.
4. Практическое применение инструментов интеграции.