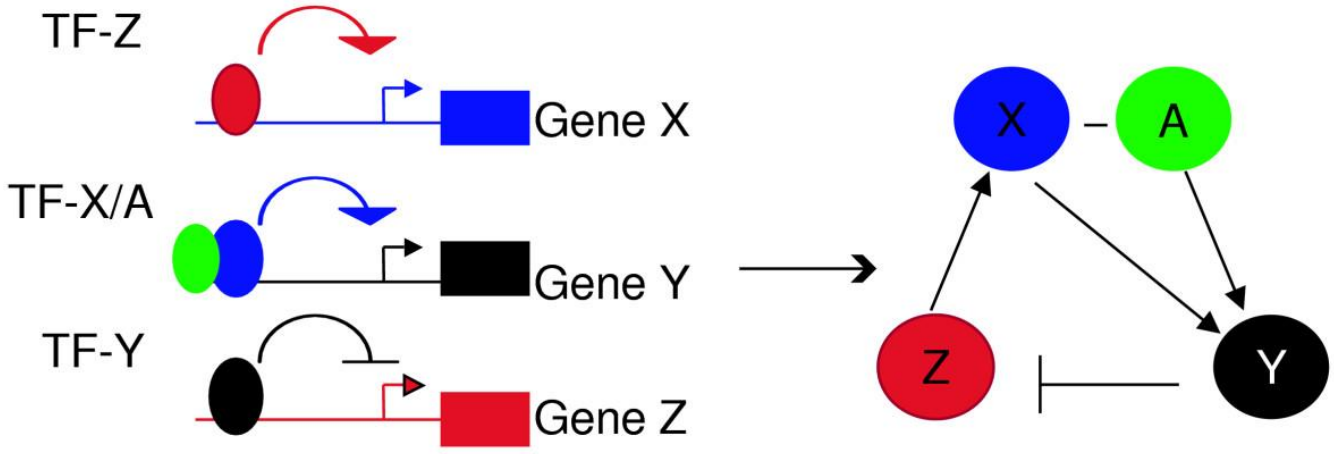
Лекция №4

1. Генные сети в эпоху высокопроизводительного секвенирования

2. Базы данных по генным сетям

к.б.н., с.н.с. лаб. эволюционной биоинформатики
и теоретической генетики Игнатьева Е.В.

Транскрипционные регуляторные сети (Transcription regulatory networks)



Транскрипционные регуляторные сети содержат узлы, обозначающие одновременно два вида объектов:

1) гены, относящиеся к рассматриваемой системе;

2) транскрипционные факторы, регулирующие экспрессию генов.

Типы связей (ребер) в транскрипционных регуляторных сетях:

Стрелки отображают активацию (подавление) экспрессии гена транскрипционным фактором,

Линии отображают белок-белковые взаимодействия

При отображении транскрипционных регуляторных сетей не принято изображать отдельно ген и отдельно белок !!!

Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA, Walhout AJ. A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol. 2005;6(13):R110.

Итоги: возможные экспериментальные подходы для получения данных для реконструкции транскрипционных регуляторных сетей (TRN)

√ Подход 1. Выявление сайтов чувствительности к ДНКазе-I в геноме и предсказание сайтов связывания транскрипционных факторов в районах ДНК, соответствующих сайтам чувствительности к ДНКазе-I

Подход 2. Оценка эффекта выключения активности транскрипционных факторов различными способами (siRNA и т.д.).

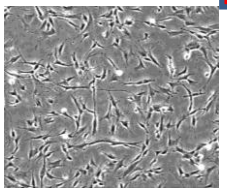
. . . . . .

. . . . . .

Подход 2: Реконструкция транскрипционных регуляторных сетей (TRN) на основе исследования эффектов выключения генов с помощью siRNA (метод)

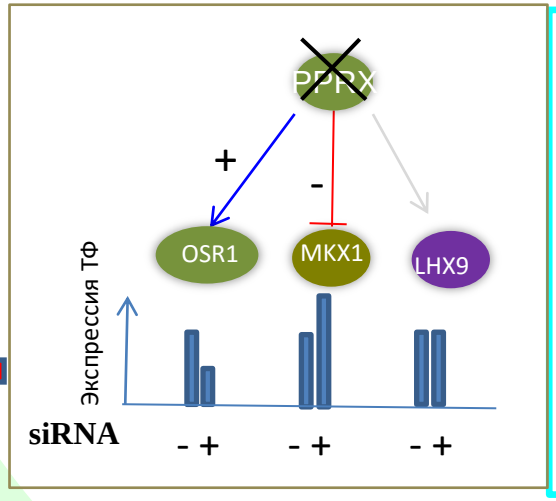
Анализ данных консорциума FANTOM5 по экспрессии в различных типах клеток человека (методика CAGE)

Список 18 транскрип-ционных факторов с повышенной экспрессией в фибробластах (fibroblast-enriched transcription factors)



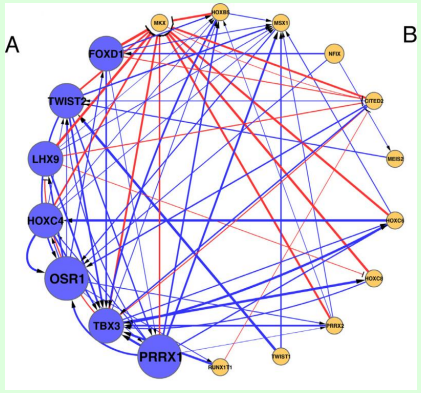
Эксперимент с поочередным выключением каждого из 18-ти ТФ с помощью siRNA

siRNA (= малые интерферирующие РНК) могут быть искусственно введены в клетки для нокдауна определённого гена.



Выявление генов, экспрессия которых меняется наиболее сильно в ответ на вмешательство siRNA .
= Это гены мишени, регулируемые исследуемыми ТФ

Фибробласт-специфичная сеть транскрипционной регуляции (TRN)

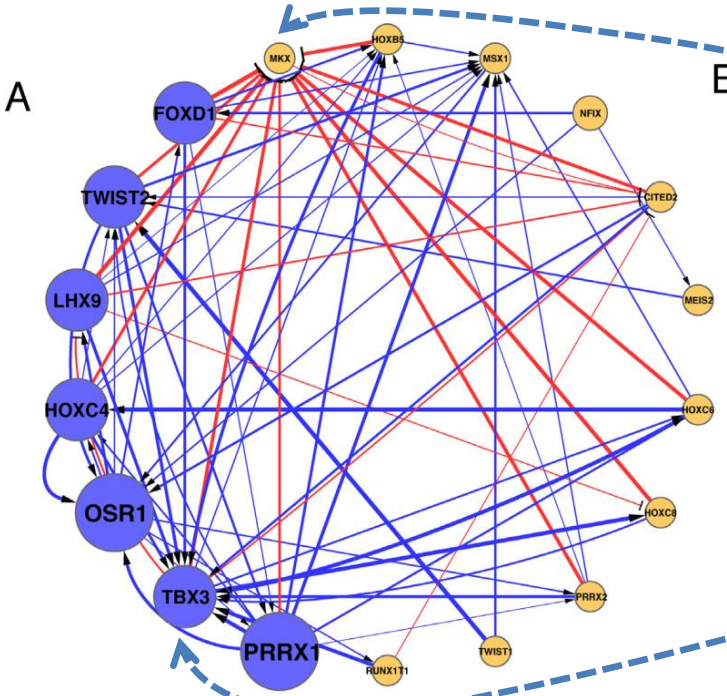


7 транскрипционных факторов с наибольшей значимостью + 11 других факторов

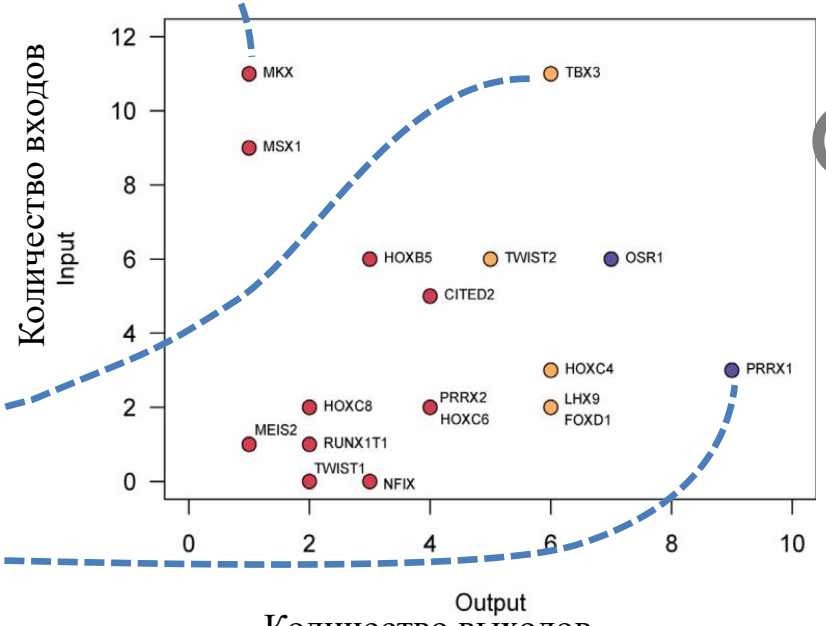
Tomaru Y, et al + FANTOM Consortium, Shin JW, Suzuki H. A transient disruption of fibroblastic transcriptional regulatory network facilitates trans-differentiation. Nucleic Acids Res. 2014 Aug;42(14):8905-13

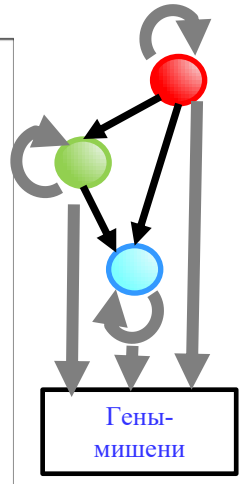
Подход 2: Реконструкция транскрипционных регуляторных сетей (TRN) на основе исследования эффектов выключения генов с помощью siRNA (результаты)

Фибробласт-специфичная сеть транскрипционной регуляции (TRN)



Характеристика узлов в сети по количеству входящих и выходящих регуляторных влияний



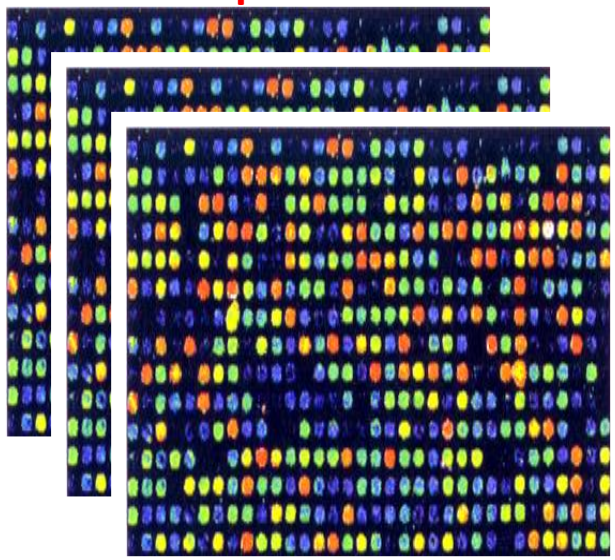


7 транскрипционных факторов с наибольшей значимостью + 11 других факторов

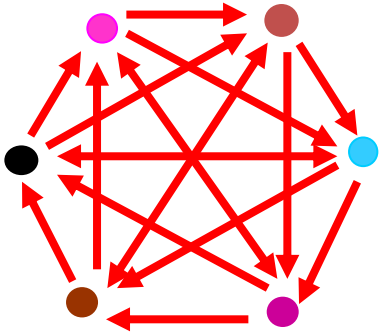
Tomaru Y, et al + FANTOM Consortium, Shin JW, Suzuki H. A transient disruption of fibroblastic transcriptional regulatory network facilitates trans-differentiation. Nucleic Acids Res. 2014 Aug;42(14):8905-13

Задача в общем виде:
Реконструкция графов регуляторных генных сетей на основе данных по экспрессионным ДНК-чипам

Дано: Экспериментальные данные по уровням экспрессии генов в нескольких временных точках



Надо построить:

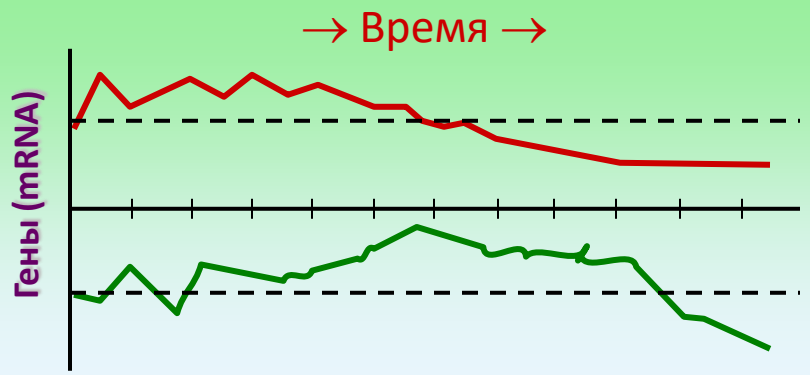


Граф взаимодействий между генами

I ⇔ концентрация *i*-го гена в *j*-й временной точке

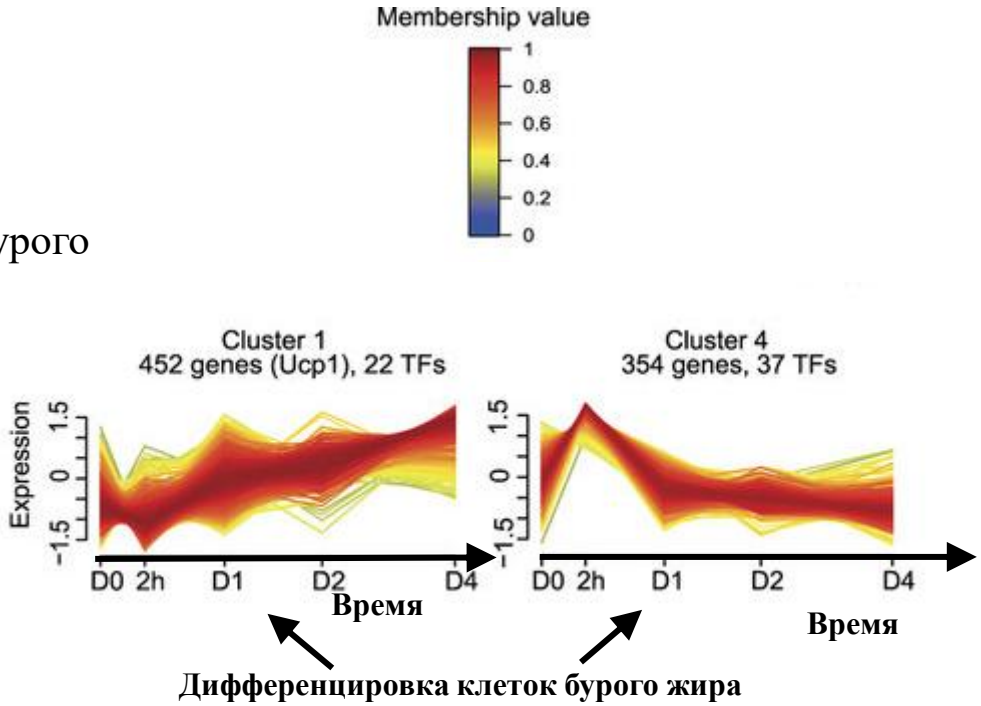
→ Время →

$$X_{N \times M} := \begin{pmatrix} \chi_1^1 & \chi_1^2 & \cdots & \chi_1^M \\ \chi_2^1 & \chi_2^2 & \cdots & \chi_2^M \\ \vdots & \vdots & \ddots & \vdots \\ \chi_N^1 & \chi_N^2 & \cdots & \chi_N^M \end{pmatrix}$$



Один из этапов теоретического подхода: выявление кластеров генов с контрастными экспрессионными характеристиками

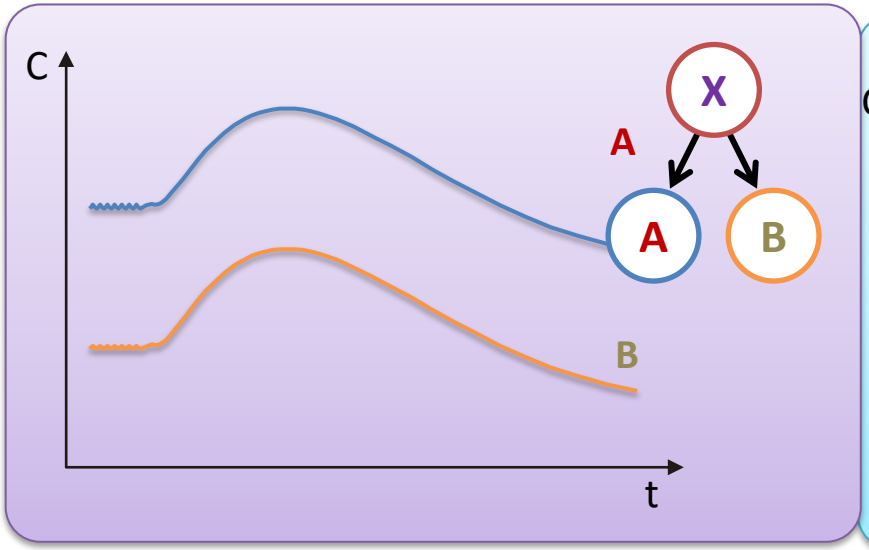
ГО термины:
Дифференцировка бурого жира
Клеточный ответ на инсулин
Биосинтез жирных кислот
…..

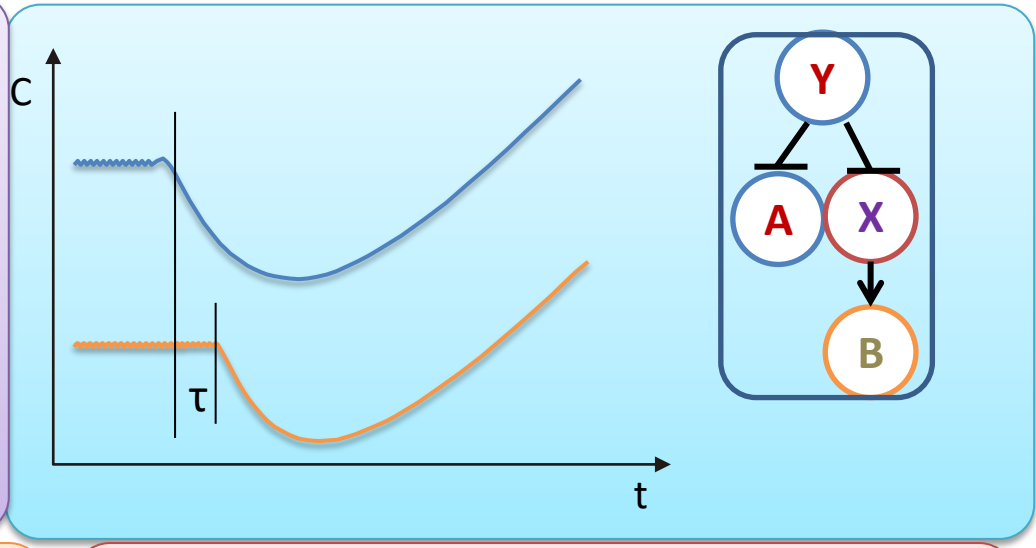


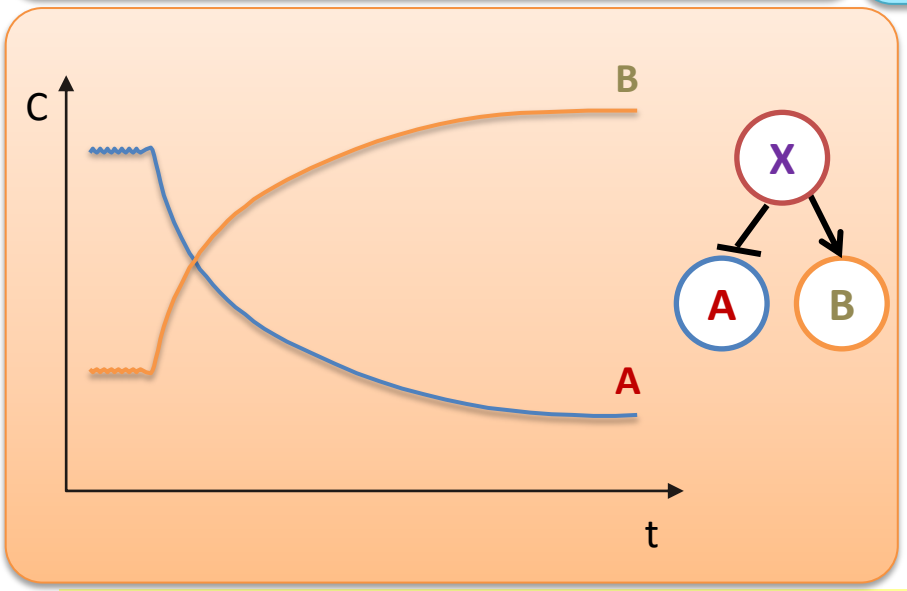
ГО термины:
Морфогенез эпителиальной трубки
Пролиферация клеток
Дифференцировка хондроцитов
…..

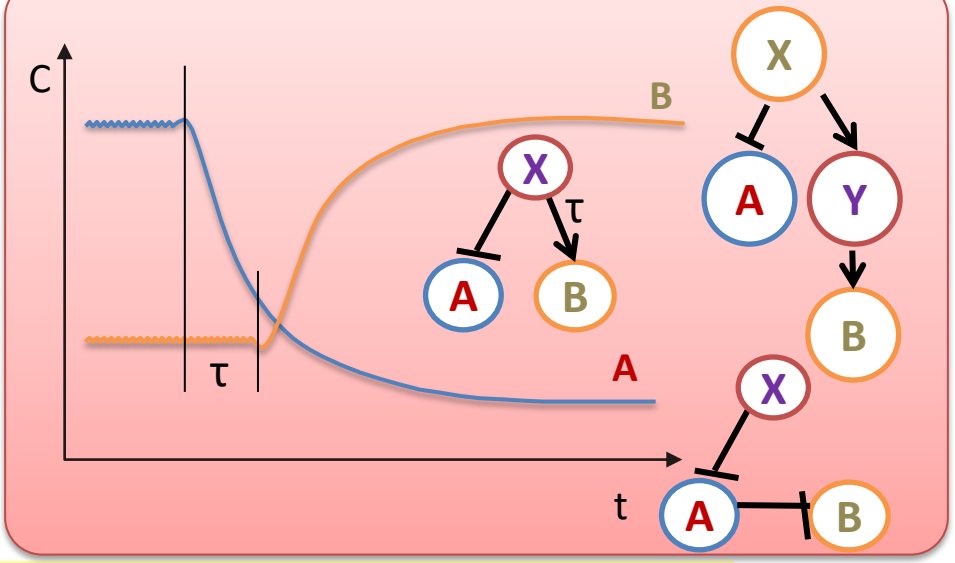
Pradhan RN et al. Dissecting the brown adipogenic regulatory network using integrative genomics. Sci Rep. 2017 Feb 9;7:42130.

Построение гипотез о регуляторных взаимодействиях между генами основе данных по экспрессионным ДНК-чипам





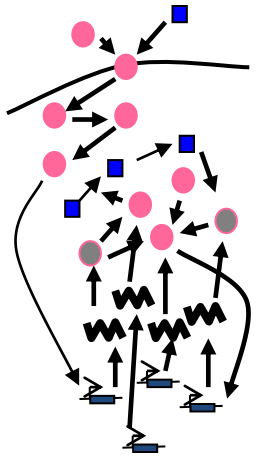




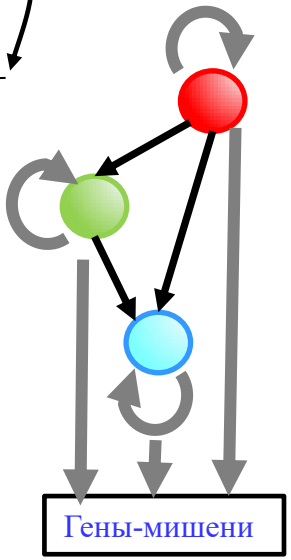
Результаты неоднозначны, необходима экспериментальная проверка !!!!

Колчанов Н.А., и др. ГЕННЫЕ СЕТИ. Вавиловский журнал генетики и селекции, 2013, Том 17, № 4/2, стр.833-850

Генные сети, транскрипционные регуляторные сети- Что дальше ????



√ **Генные сети** - молекулярно-генетические системы, обеспечивающие формирование фенотипических характеристик организмов (молекулярных, биохимических, структурных, морфологических, поведенческих и т.д.) на основе информации, закодированных в их геномах.



√ **Транскрипционные регуляторные сети (TRN)** – сети, содержащие вершины, каждая из которых соответствует двум видам объектов:
1) гены, относящиеся к рассматриваемой системе;
2) транскрипционные факторы, кодируемые данными генами, и регулирующие экспрессию других генов из данной системы.

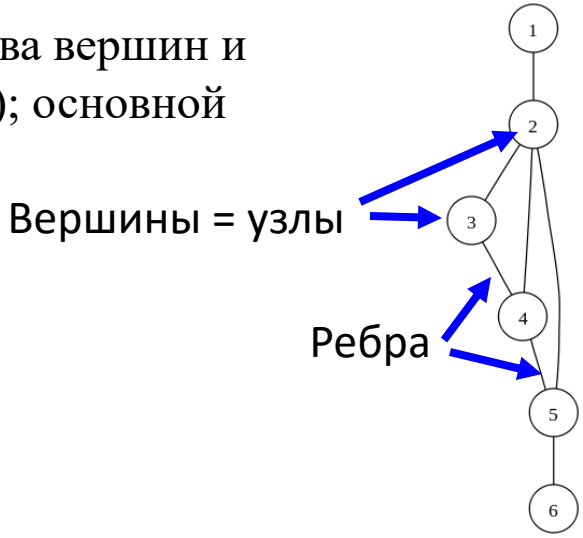
Сети взаимодействий между генами / белками

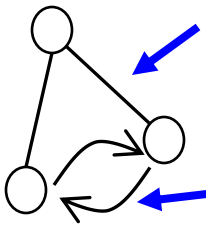
. . . . . .

Изображения генных сетей, а также любых сетей, являются графами.

Граф (англ. *graph*) — совокупность непустого множества вершин и наборов пар вершин (связей между вершинами = ребер); основной объект изучения математической теории графов.

https://ru.wikipedia.org/wiki/ Граф_(математика)



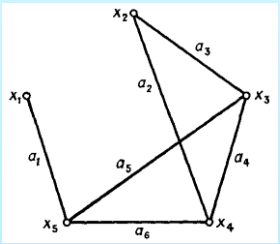


Неориентированное ребро соединяет неупорядоченную пару вершин

Ориентированное ребро (= дуга) соединяет упорядоченную пару вершин

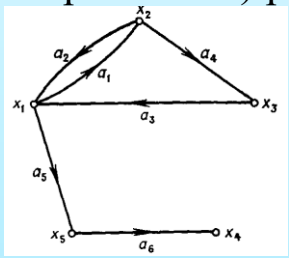
Классификация графов по типам входящих в них ребер

Неориентированный граф — это непустое множество вершин (=узлов) и неупорядоченных ребер



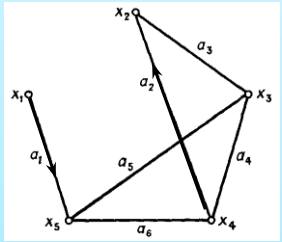
Пример неориентированного графа. x1, x2, x3, x4, x5 – вершины графа. a1, a2, a3, a4, a5, a6 – ребра.

Ориентированный граф — это непустое множество вершин (=узлов) и упорядоченных (ориентированных) ребер (дуг)



Пример ориентированного графа. x_1, x_2, x_3, x_4, x_5 – вершины графа. a_1, a_2, a_3, a_4, a_5, a_6 – дуги.

Смешанный граф это граф, в котором некоторые рёбра могут быть ориен-тированными, а некоторые - неориентированными.

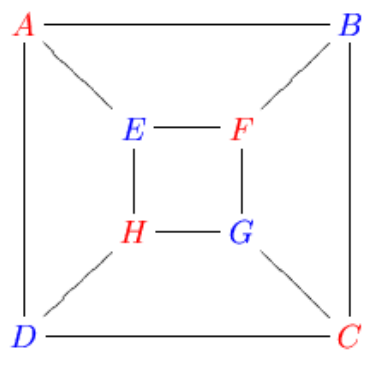


Пример смешанного графа. x_1, x_2, x_3, x_4, x_5 – вершины графа. a_1, a_2 – дуги. a_3, a_4, a_5, a_6 – ребра.

Классификация графов по типам входящих в них ребер

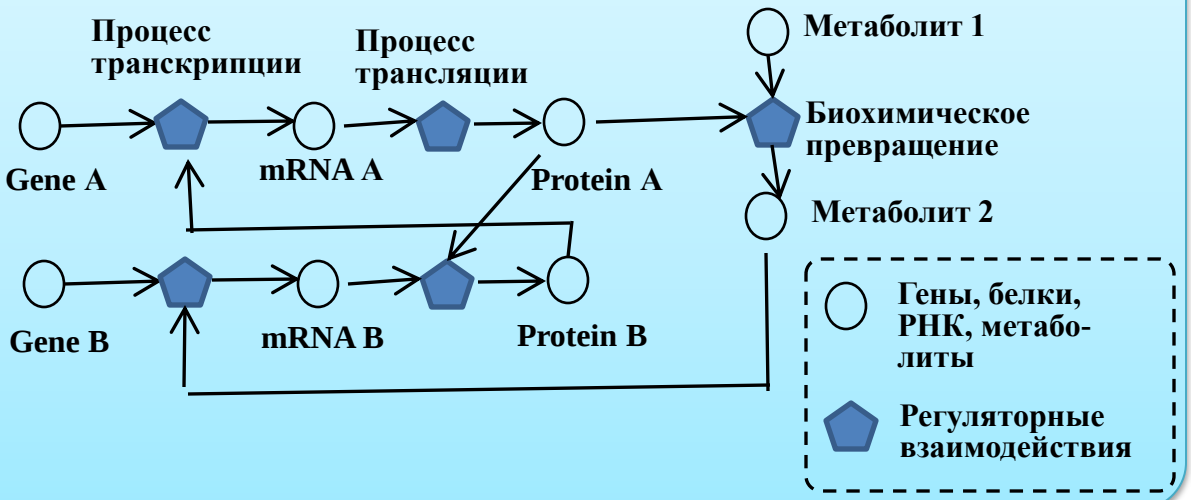
Граф называется двудольным, если его вершины можно разбить на два непересекающихся подмножества V1 и V2 так, что всякое ребро соединяет вершину из первого множества V1 с вершиной из второго множества V2.

Граф называется k-дольным, если его вершины можно разбить на **k** непересекающихся подмножества V1, V2, …, Vk, так, так, что не будет рёбер, соединяющих вершины одного и того же подмножества

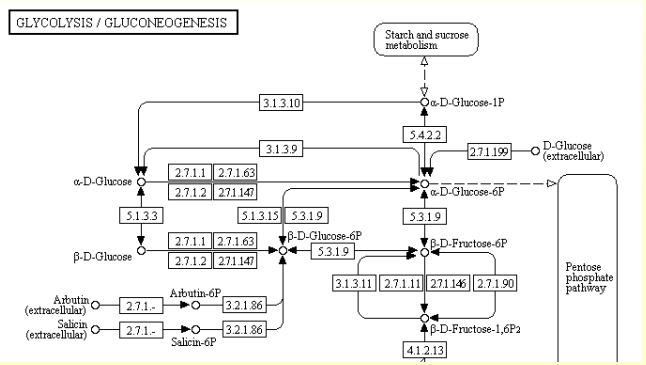


Пример друдольного графа. Первое множество вершин (**A, F, H, C**) и второе множество вершин (**E, B, D, G**).

Генную сеть можно представить в виде двудольного графа:
1-ый тип вершин – гены, белки, РНК, метаболиты
2-ой тип вершин – процессы и регуляторные взаимодействия

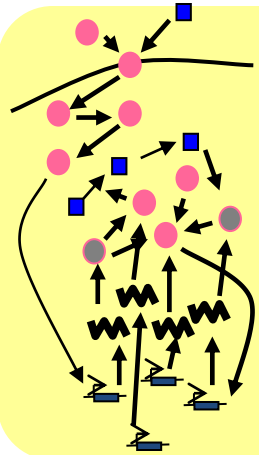


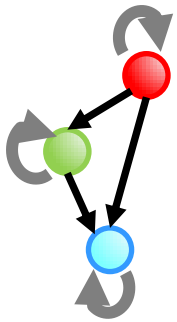
Представление данных о метаболическом пути в базе KEGG:
1-ый тип вершин – метаболиты
2-ой тип вершин – ферменты



Классификация сетей по типам вершин и связей

Генные сети – граф сети **ориентированный либо смешанный** и может включать **вершины нескольких типов**: гены, мРНК, белки, низкомолекулярные соединения, реакции





Транскрипционные регуляторные сети (TRN) граф сети **ориентированный** и включает **вершины одного типа**, обозначающие (одновременно !!!) : 1) гены, относящиеся к рассматриваемой системе; 2) транскрипционные факторы, кодируемые данными генами и регулирующие экспрессию генов

NEW !!!

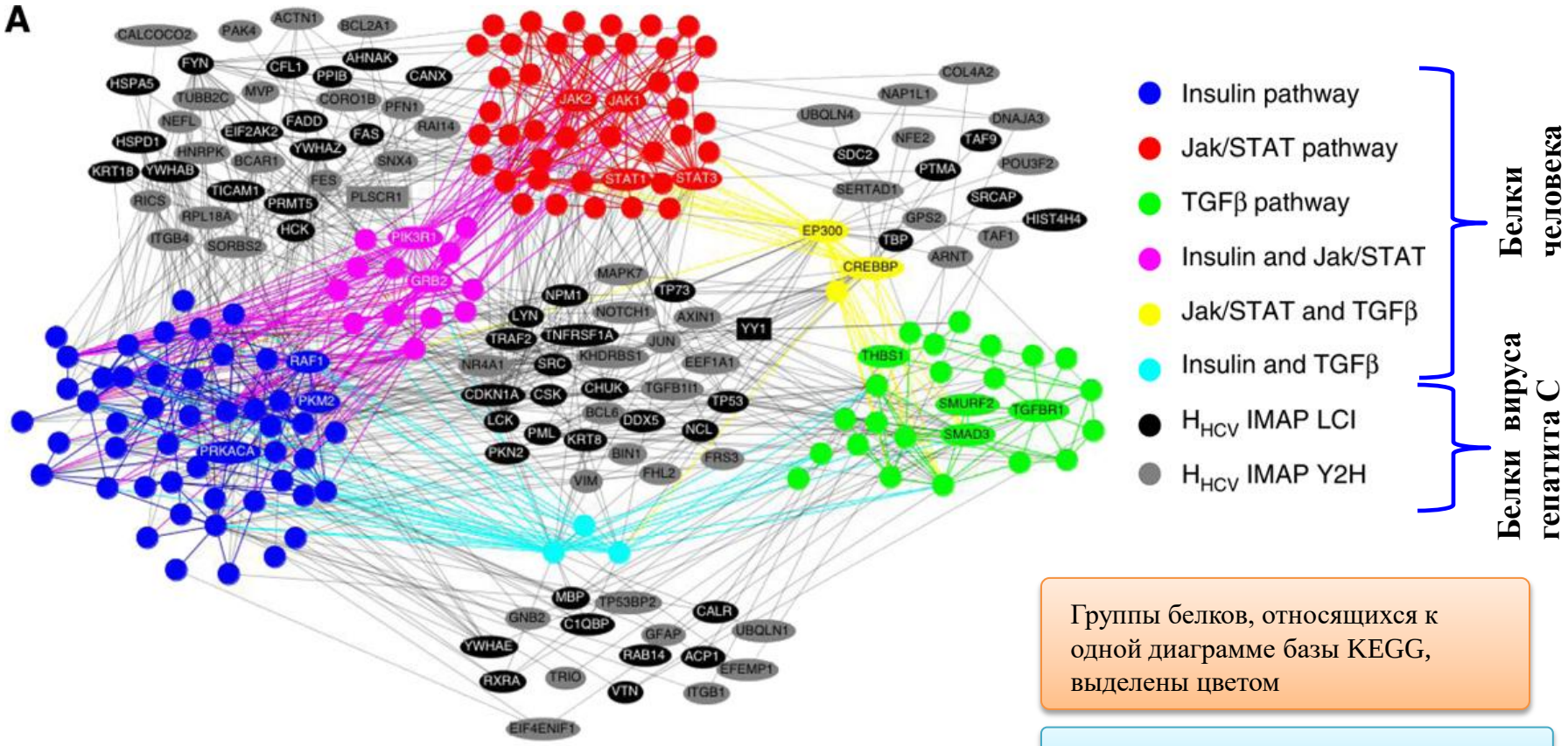
Сети взаимодействий между генами / белками - граф сети **неориентированный либо смешанный,** включает **вершины, одного типа**, обозначающие гены и /или белки.

Ребро (связь) между вершинами сети может обозначать:

- **Белок-белковое взаимо-действие**
- **Гомологию между генами либо структурное сходство белков**
- **Коэкспрессию**
- **Принадлежность генов/белков к одному метаболическому-сигнальному пути**
- **Совместную встречаемость в текстах статей**

Сети взаимодействий между генами / белками широко используются для представления данных полногеномных исследований

Пример №1: представление сети белок-белковых взаимодействий между белками человека, имеющими взаимодействия с белками вируса гепатита С



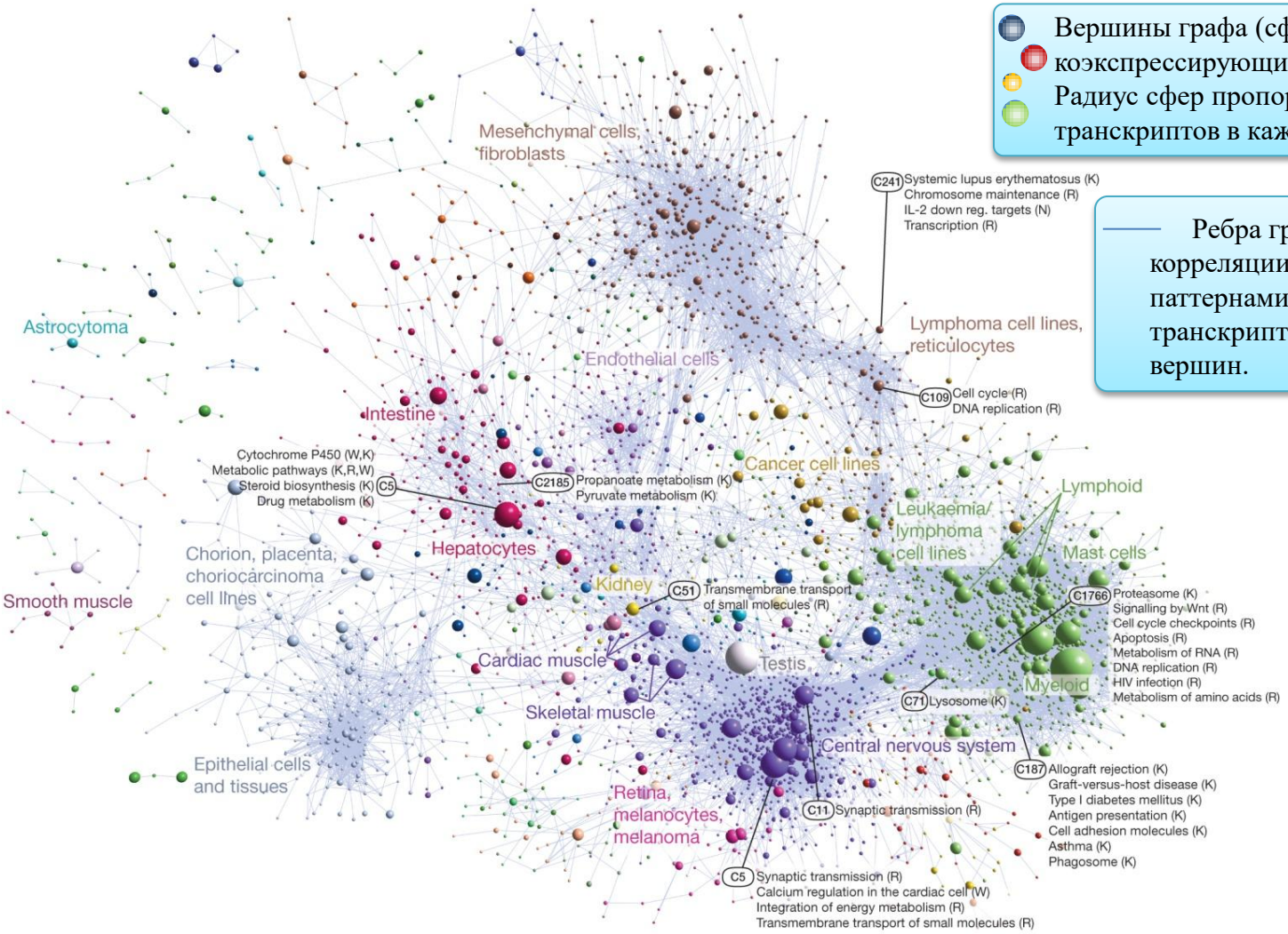
Группы белков, относящихся к одной диаграмме базы KEGG, выделены цветом

Ребра графа обозначают белок-белковые взаимодействия

B de Chassey et al. Hepatitis C virus infection protein network.// Molecular Systems Biology (2008) 4, 230.

Сети взаимодействий между генами / белками широко используются для представления данных полногеномных исследований

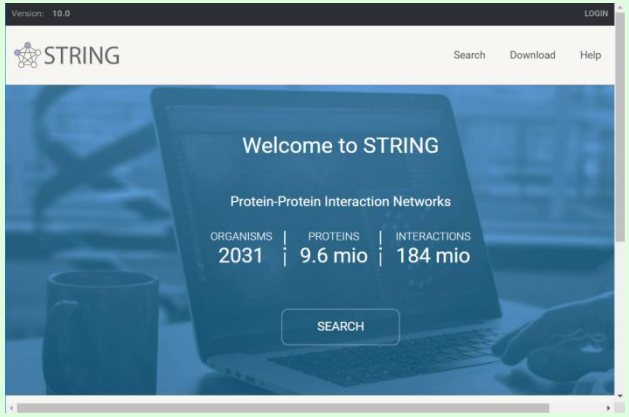
Пример №2: представление коэкспрессирующихся генов/транскриптов человека на основе анализа паттернов экспрессии , полученных консорциумом FANTOM



FANTOM Consortium and the RIKEN PMI and CLST (DGT), et al. Promoter-level mammalian expression atlas.// Nature. 2014. V. 507. P. 462-470.

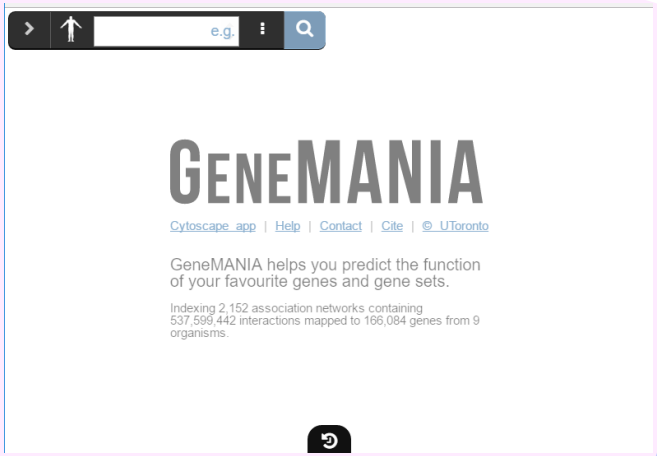
Как строить сети взаимодействий между генами / белками ??

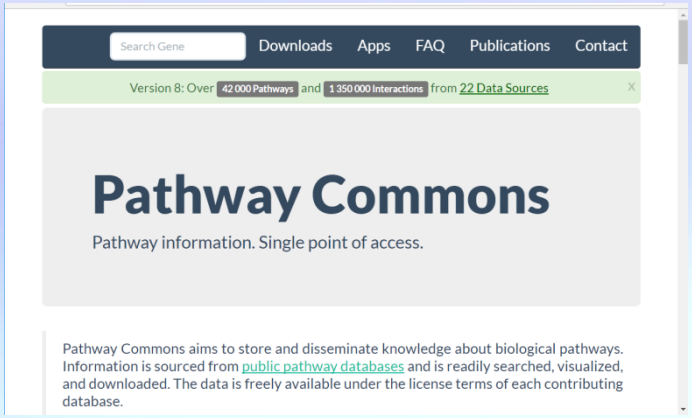
Интернет-доступные информационные компьютерные системы, позволяющие экстрагировать данные по связям различных типов между генами/белками



STRING - functional protein association networks
string-db.org/

GeneMANIA
http://genemania.org/

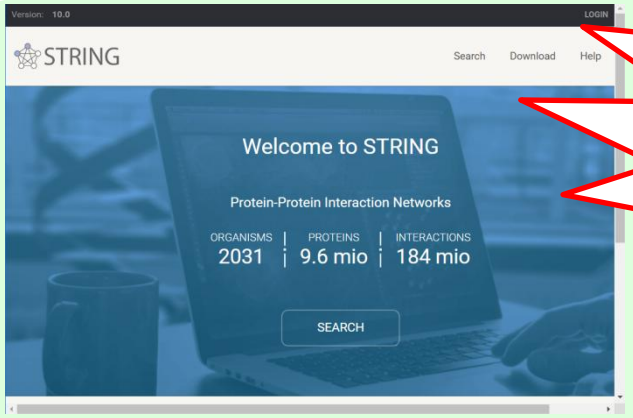




Pathway Commons - A Resource for Biological Pathway Analysis
http://www.pathwaycommons.org/

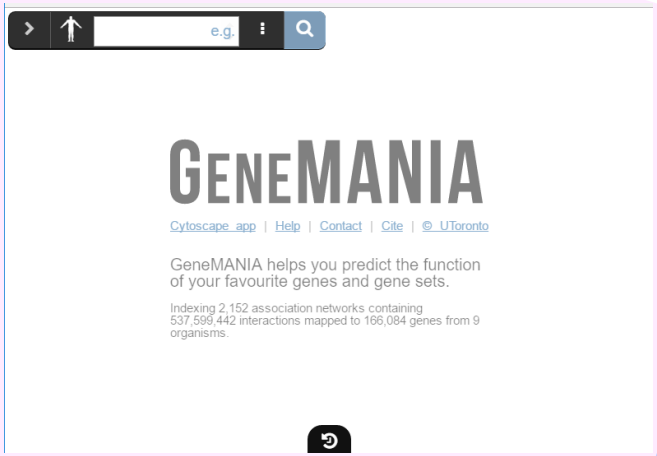
Как строить сети взаимодействий между генами / белками ??

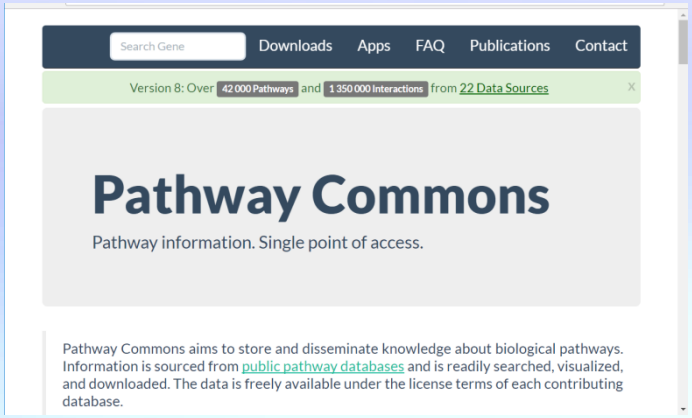
Интернет-доступные информационные компьютерные системы, позволяющие экстрагировать данные по связям различных типов между генами/белками



STRING - functional protein association networks
string-db.org/

GeneMANIA
http://genemania.org/

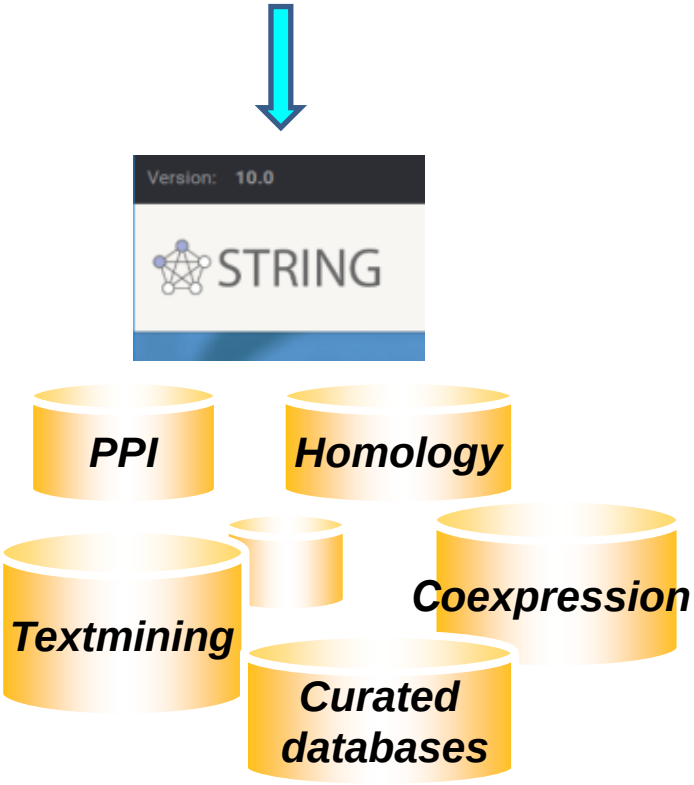




Pathway Commons - A Resource for Biological Pathway Analysis
http://www.pathwaycommons.org/

Программно-информационный инструмент STRING
(http://string-db.org/)

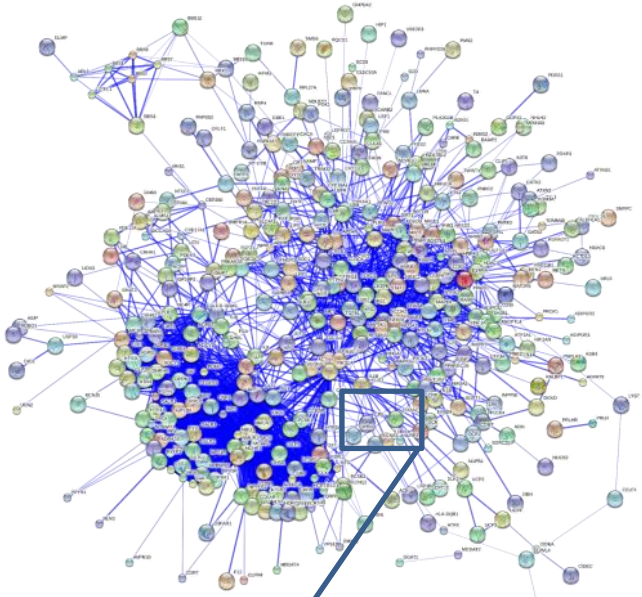
Дано: список ~500 генов человека



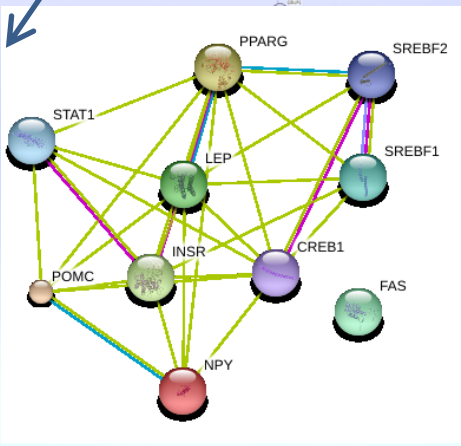
Базы данных взаимодействий между генами

849 видов организмов с пометкой “*core*”
1184 видов организмов с пометкой “*periphery*”

Общий вид сети взаимодействий между генами, вовлеченными в регуляцию массы тела

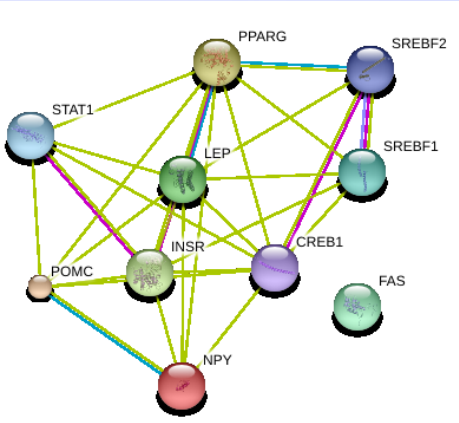


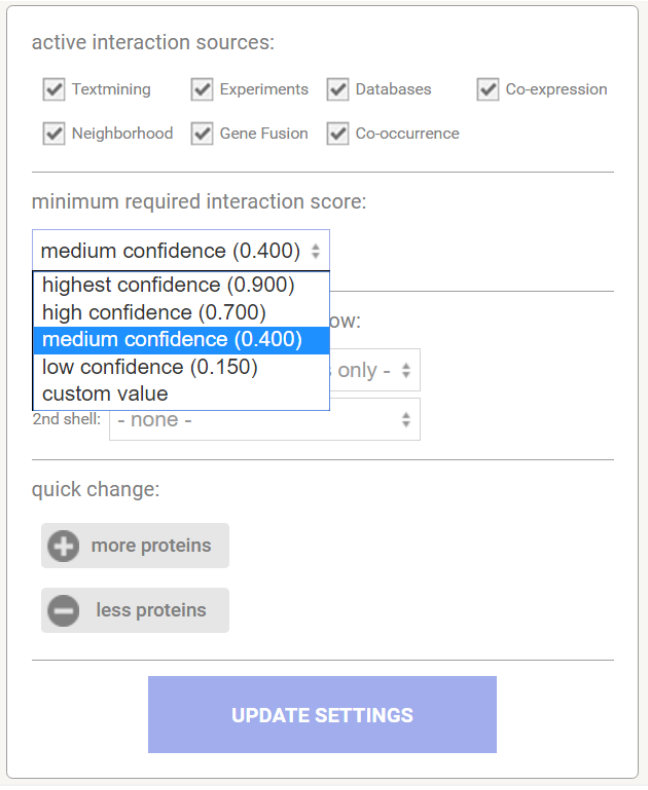
Фрагмент сети



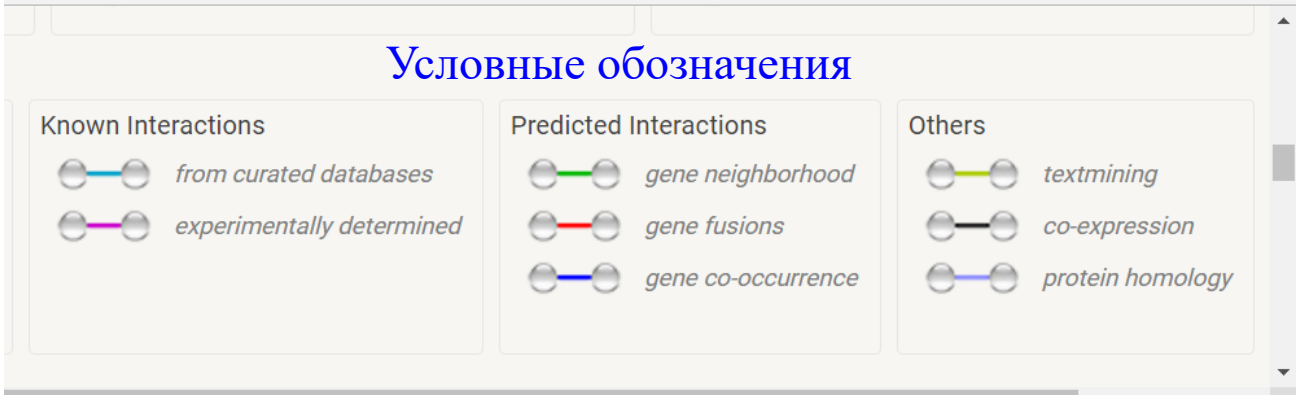
Сеть, реконструированная системой STRING, включает взаимодействия разных типов

Фрагмент сети

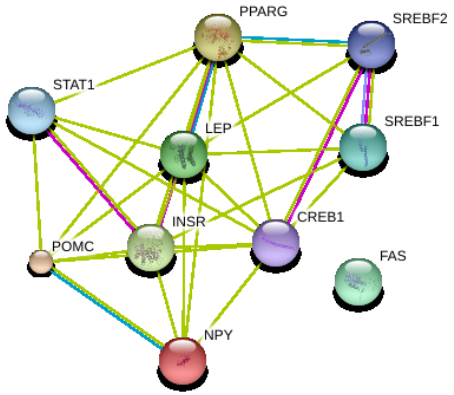


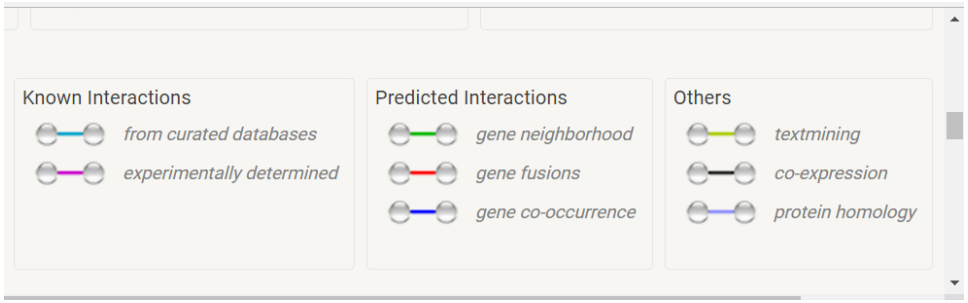


Условные обозначения



Основные типы взаимодействий в системе STRING





***Experimentally determined* - белок-белковые взаимодействия из баз данных (базы, входящие в IMEx consortium (HPRD, MINT, etc.), а также BioGRID)**

***From curated databases* - совместная встречаемость в метаболических и сигнальных путях**

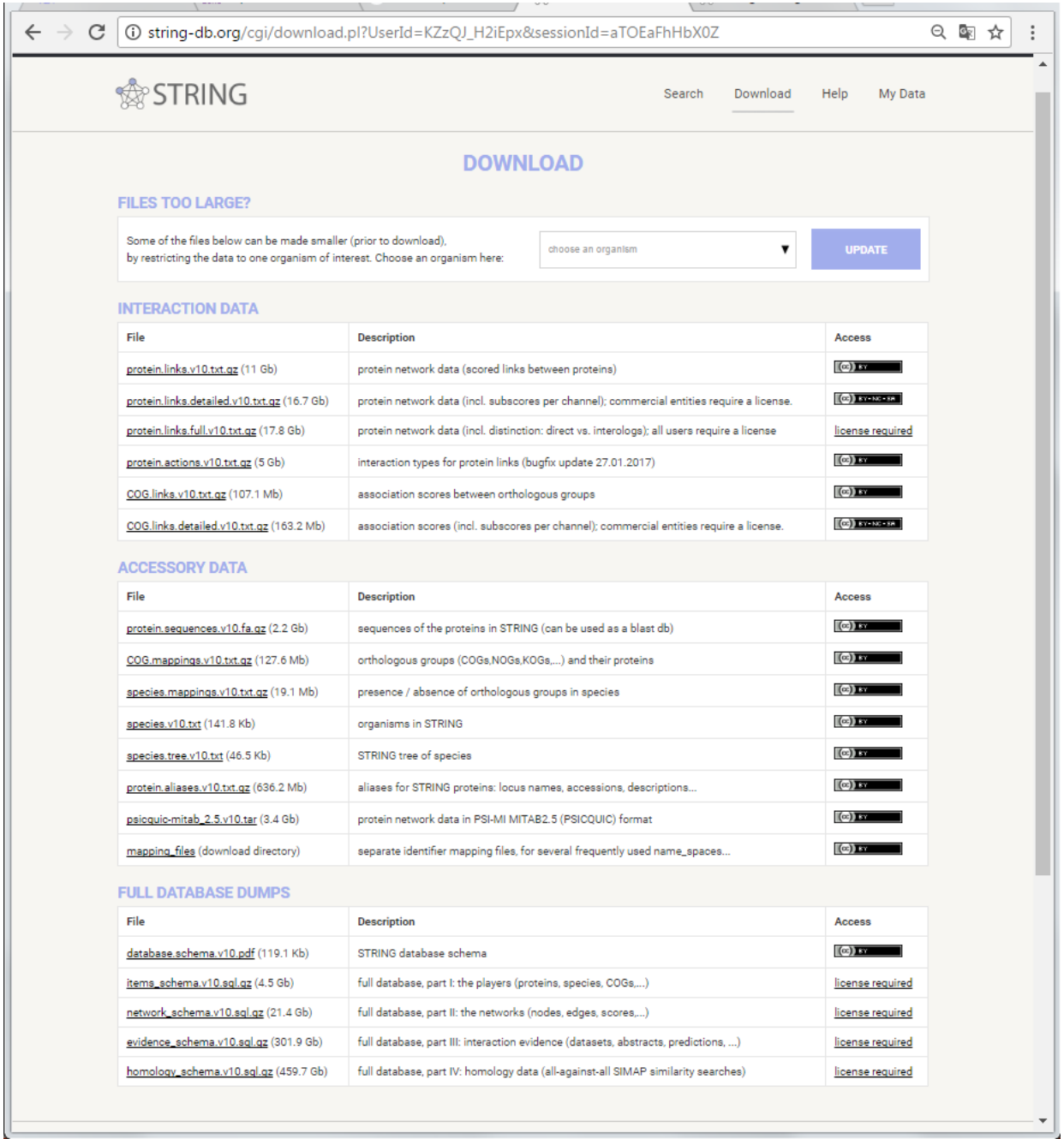
***Textmining* – встречаемость в абстрактах и текстах статей**

***Coexpression* – сходные паттерны экспрессии генов**

***Protein homology* – структурное сходство белков (например, семейства белков)**

Szklarczyk D et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible.Nucleic Acids Res. 2017 Jan 4;45(D1):D362-D368.

STRING: данные можно скачать в тестовом виде !!



string-db.org/cgi/download.pl?UserId=KZzQJ_H2iEpx&sessionId=aTOEaFhHbX0Z

STRING

Search | Download | Help | My Data

DOWNLOAD

FILES TOO LARGE?

Some of the files below can be made smaller (prior to download), by restricting the data to one organism of interest. Choose an organism here: choose an organism UPDATE

INTERACTION DATA

File	Description	Access
protein.links.v10.txt.gz (11 Gb)	protein network data (scored links between proteins)	
protein.links.detailed.v10.txt.gz (16.7 Gb)	protein network data (incl. subscores per channel); commercial entities require a license.	
protein.links.full.v10.txt.gz (17.8 Gb)	protein network data (incl. distinction: direct vs. interologs); all users require a license	license required
protein.actions.v10.txt.gz (5 Gb)	interaction types for protein links (bugfix update 27.01.2017)	
COG.links.v10.txt.gz (107.1 Mb)	association scores between orthologous groups	
COG.links.detailed.v10.txt.gz (163.2 Mb)	association scores (incl. subscores per channel); commercial entities require a license.	

ACCESSORY DATA

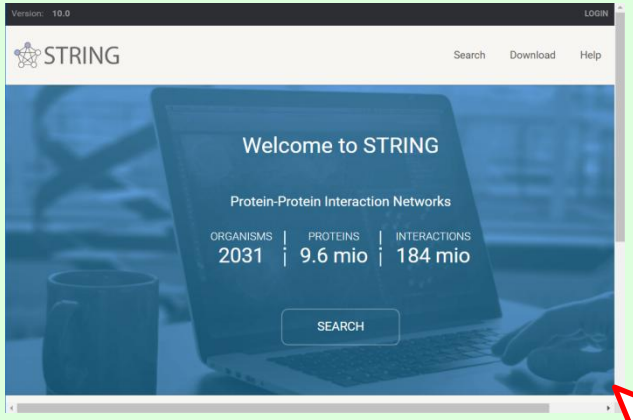
File	Description	Access
protein.sequences.v10.fa.gz (2.2 Gb)	sequences of the proteins in STRING (can be used as a blast db)	
COG.mappings.v10.txt.gz (127.6 Mb)	orthologous groups (COGs,NOGs,KOGs,...) and their proteins	
species.mappings.v10.txt.gz (19.1 Mb)	presence / absence of orthologous groups in species	
species.v10.txt (141.8 Kb)	organisms in STRING	
species.tree.v10.txt (46.5 Kb)	STRING tree of species	
protein.aliases.v10.txt.gz (636.2 Mb)	aliases for STRING proteins: locus names, accessions, descriptions...	
psicquic-mitab_2.5.v10.tar (3.4 Gb)	protein network data in PSI-MI MITAB2.5 (PSICQUIC) format	
mapping_files (download directory)	separate identifier mapping files, for several frequently used name_spaces...	

FULL DATABASE DUMPS

File	Description	Access
database.schema.v10.pdf (119.1 Kb)	STRING database schema	
items_schema.v10.sql.gz (4.5 Gb)	full database, part I: the players (proteins, species, COGs,...)	license required
network_schema.v10.sql.gz (21.4 Gb)	full database, part II: the networks (nodes, edges, scores,...)	license required
evidence_schema.v10.sql.gz (301.9 Gb)	full database, part III: interaction evidence (datasets, abstracts, predictions, ...)	license required
homology_schema.v10.sql.gz (459.7 Gb)	full database, part IV: homology data (all-against-all SIMAP similarity searches)	license required

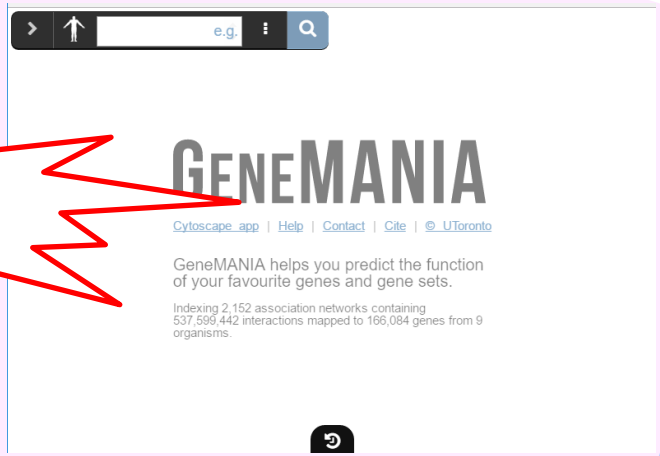
Как строить сети взаимодействий между генами / белками ??

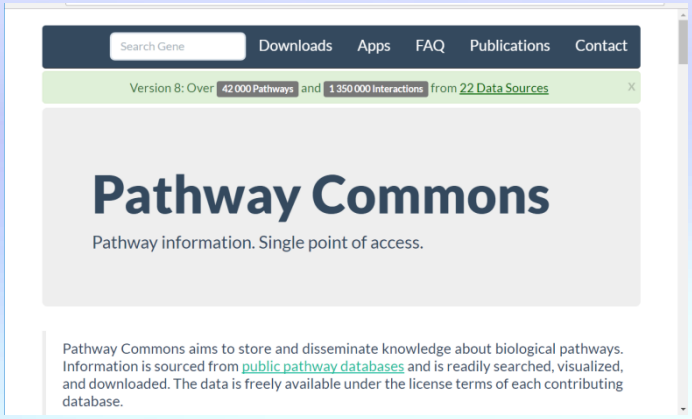
Интернет-доступные информационные компьютерные системы, позволяющие экстрагировать данные по связям различных типов между генами/белками



STRING - functional protein association networks
string-db.org/

GeneMANIA
http://genemania.org/

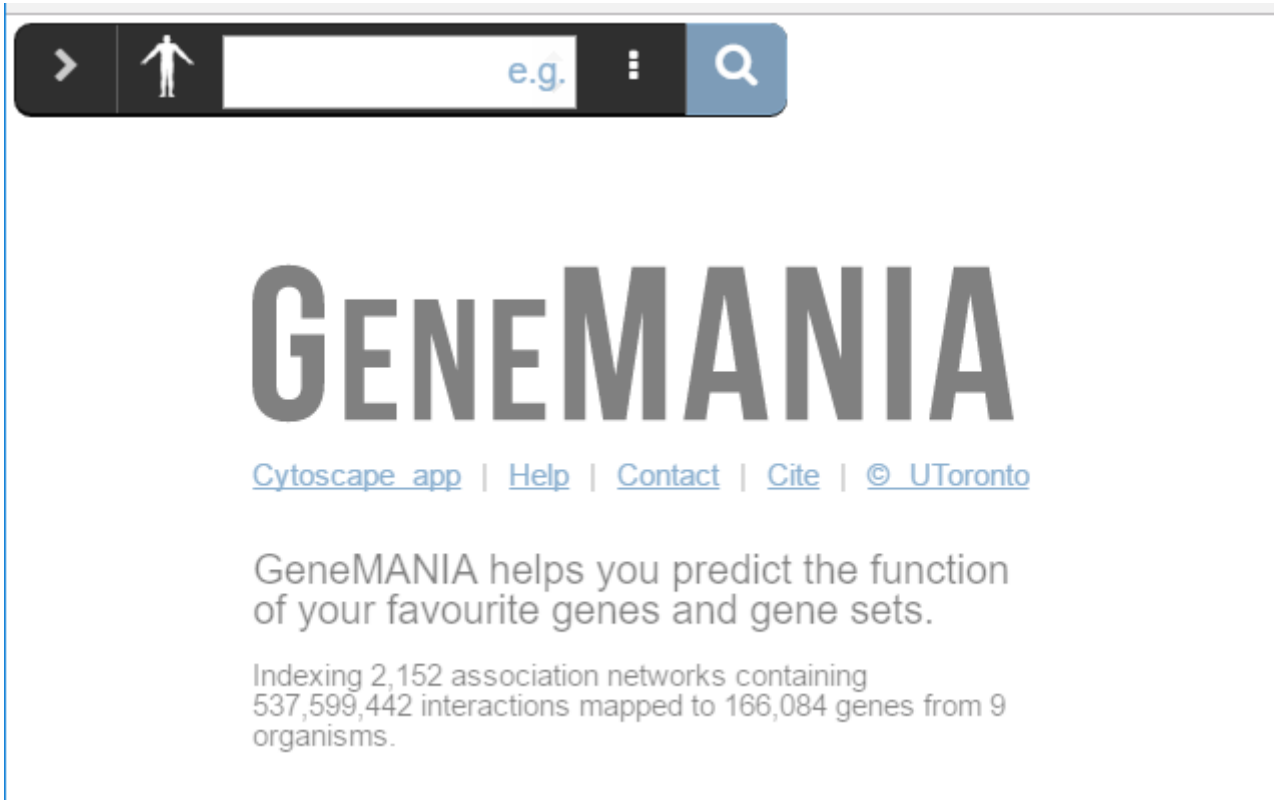




Pathway Commons - A Resource for Biological Pathway Analysis
http://www.pathwaycommons.org/

GeneMania

http://genemania.org/



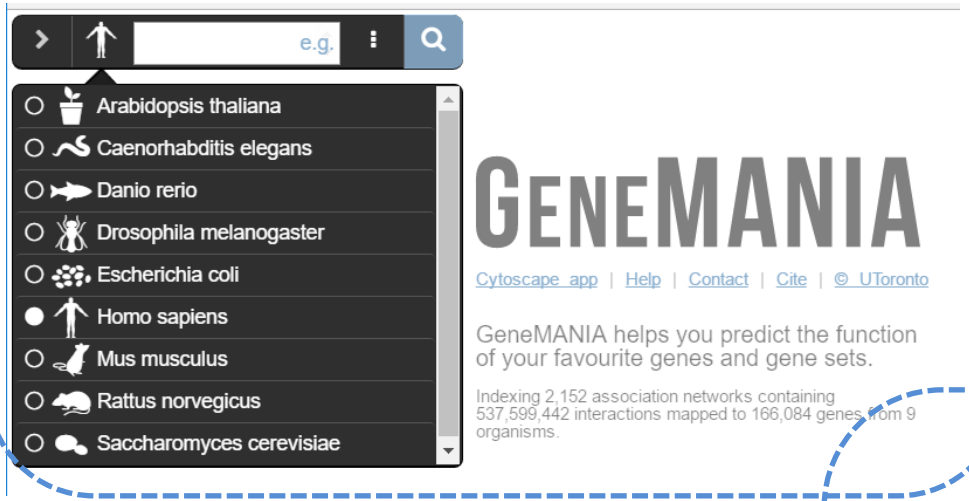
Ресурс, позволяющий определять функциональность генов и наборов генов. Имеет две реализации: Интернет-доступную, а также как плагин системы Cytoscape

Zuberi K, Franz M, Rodriguez H, Montojo J, Lopes CT, Bader GD, Morris Q. GeneMANIA prediction server 2013 update. Nucleic Acids Res. 2013 Jul;41(Web Server issue):W115-22.

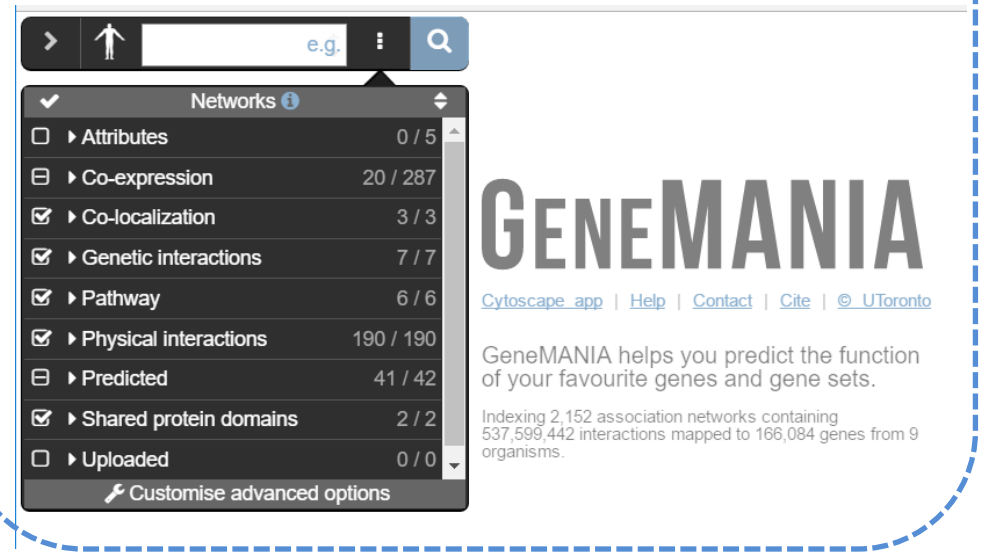
GeneMania

http://genemania.org/

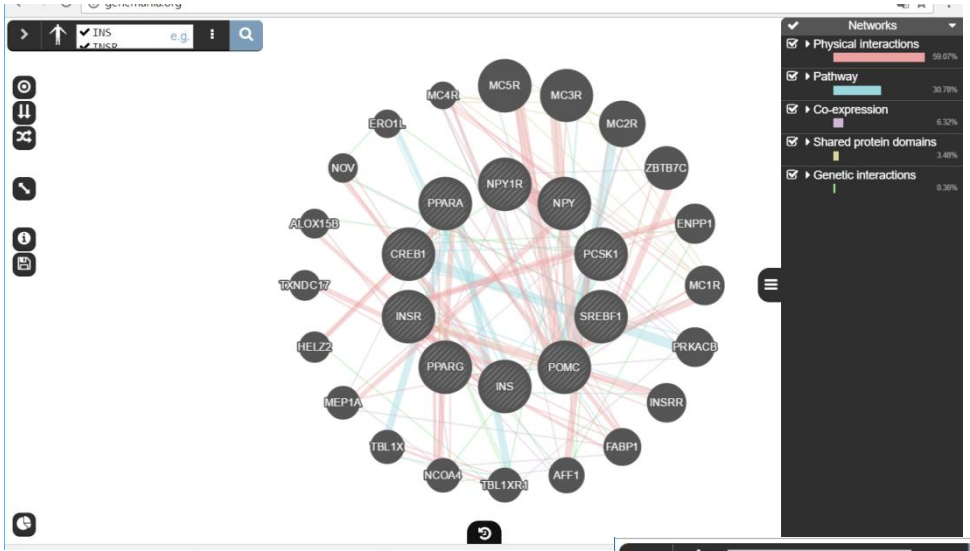
Ресурс, позволяет работать с 9-ю видами организмов



А также выбирать (комбинировать) данные различных типов:



Результат тестового запроса к GeneMania

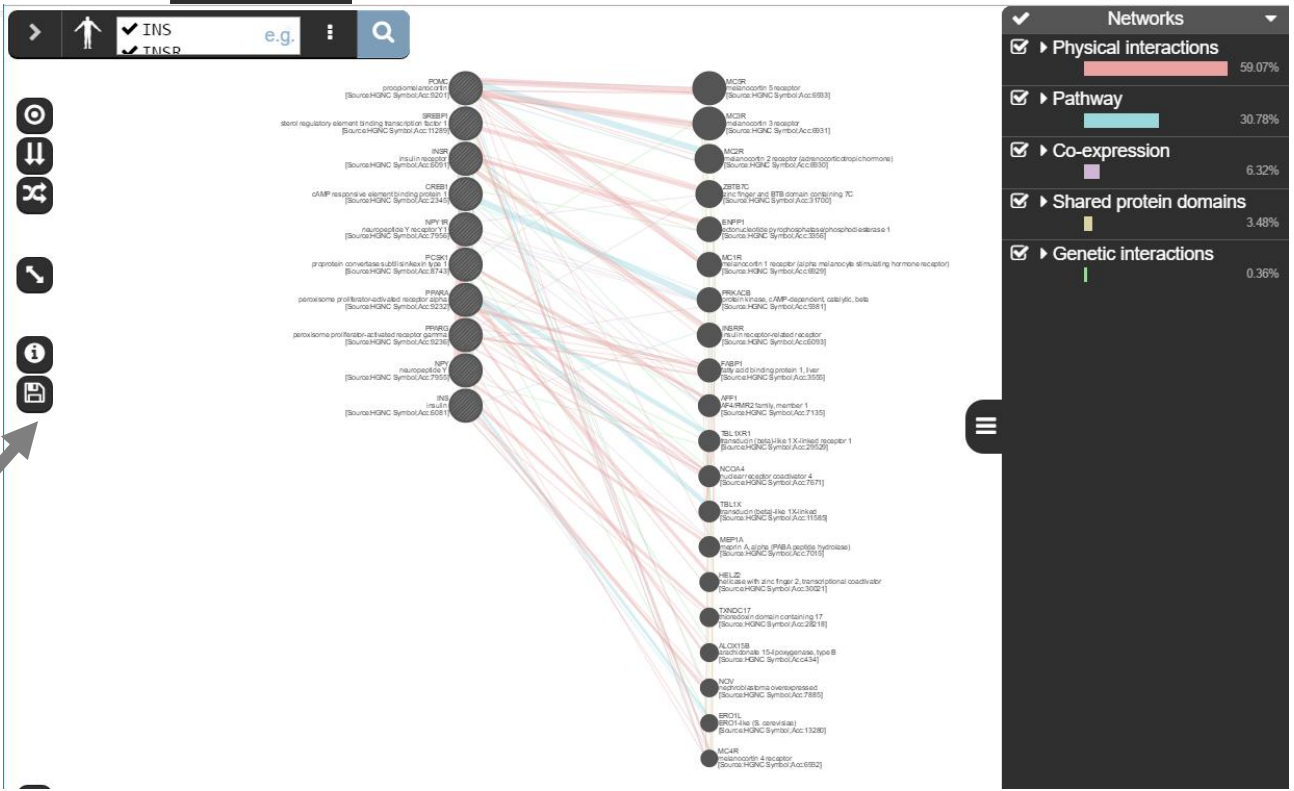


Запрос содержал 10 генов человека:
INS, INSR, POMC, SREBF1, PPARG, PPARA, NPY, NPY1R, PCSK1, CREB1

Система построила связи между 10-ю генами и дополнила список генов новыми генами, имеющими максимальное количество связей с генами из списка

Два способа представления полученного результата

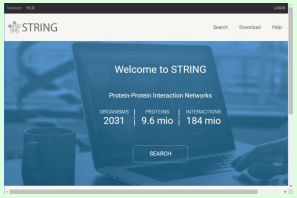
Опция для сохранения данных в текстовом виде (таблица)



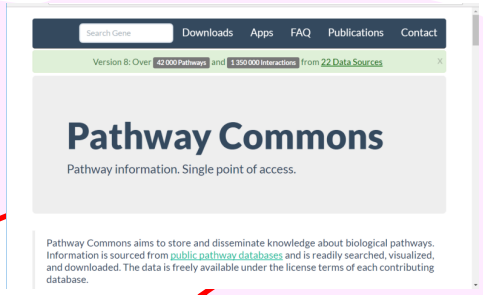
Как строить сети взаимодействий между генами / белками ??

Интернет-доступные информационные компьютерные системы, позволяющие экстрагировать данные по связям различных типов между генами/белками

STRING - functional protein association networks
string-db.org/



GeneMANIA
http://genemania.org/



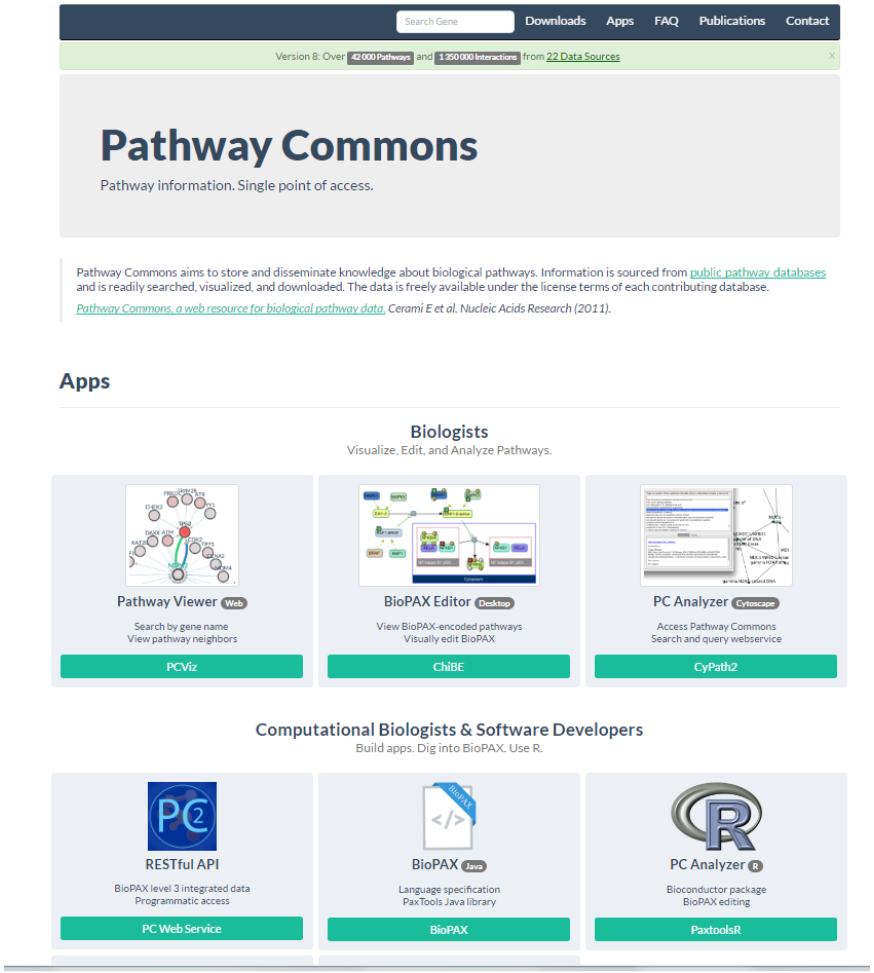
Pathway Commons
- A Resource for Biological Pathway Analysis
http://www.pathwaycommons.org/

Cytoscape **- Network Data Integration, Analysis, and Visualization in a Box**

Pathway Commons

http://www.pathwaycommons.org/

Version 8: Over **42 000 Pathways** and **1 350 000 Interactions** from 22 Data Sources



Pathway Commons aims to store and disseminate knowledge about biological pathways. Information is sourced from public pathway databases and is readily searched, visualized, and downloaded. The data is freely available under the license terms of each contributing database.

Cerami EG1, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res. 2011 Jan;39(Database issue):D685-90.

Pathway Commons: источники данных (7) по сетям

Из всех ресурсов были взяты данные, относящиеся к виду Homo sapiens

Reactome

Reactome v56 (only 'Homo sapiens.owl') 31-Mar-2016 *(BIOPAX)*
URI: http://pathwaycommons.org/pc2/reactome
All names (for data filtering): reactome
Contains: **2007** pathways, **14427** interactions, **35835** participants

NetPath

NetPath 12/2011 *(BIOPAX)*
URI: http://pathwaycommons.org/pc2/netpath
All names (for data filtering): netpath
Contains: **27** pathways, **6347** interactions, **3266** participants

NCI Pathway Interaction Database: Pathway

NCI Curated Human Pathways from PID (final); 27-Jul-2015 *(BIOPAX)*
URI: http://pathwaycommons.org/pc2/pid
All names (for data filtering): pid,nci pathway interaction database: pathway
Contains: **745** pathways, **14707** interactions, **10531** participants

KEGG Pathway

KEGG 07/2011 (only human, hsa* files), converted to BioPAX by BioModels (http://www.ebi.ac.uk/biomodels-main/) team *(BIOPAX)*
URI: http://pathwaycommons.org/pc2/kegg
All names (for data filtering): kegg,kegg pathway
Contains: **122** pathways, **3566** interactions, **3355** participants

HumanCyc

HumanCyc 19.5; 27-Oct-2015; under license from SRI International, www.biocyc.org *(BIOPAX)*
URI: http://pathwaycommons.org/pc2/humancyc
All names (for data filtering): humancyc,biocyc
Contains: **302** pathways, **7102** interactions, **5896** participants

WikiPathways

WikiPathways - Community Curated Human Pathways; 29/09/2015 (human) *(BIOPAX)*
URI: http://pathwaycommons.org/pc2/wp
All names (for data filtering): wikipathways
Contains: **333** pathways, **9758** interactions, **9584** participants

PANTHER Pathway

PANTHER Pathways 3.4 on 18-May-2015 (auto-converted to human-only model) *(BIOPAX)*
URI: http://pathwaycommons.org/pc2/panther
All names (for data filtering): panther,panther pathway,pantherdb
Contains: **272** pathways, **4700** interactions, **6703** participants

Pathway Commons : источники данных по связям и взаимодействиям в сетях

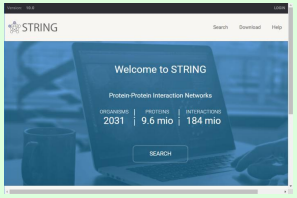
(http://www.pathwaycommons.org/pc2/datasources)

1. PhosphoSitePlus PhosphoSite Kinase-substrate information; 15-Mar-2016 (BIOPAX)
2. HPRD PSI-MI Release 9; 13-Apr-2010 (PSI_MI)
3. Database of Interacting Proteins DIP (human), 14-01-2016 (PSI_MI)
4. BioGRID Release 3.4.135 (human and the viruses), 24-Mar-2016 (PSI_MI)
5. IntAct (human only; 'negative' files removed), 16-Feb-2016 *(PSI_MI)*
6. IntAct Complex (human), 16-Feb-2016 *(PSI_MI)*
7. BIND (human), 15-Dec-2010 *(PSI_MI)*
8. CORUM (human), 17-Feb-2012 *(PSI_MI)*
9. Transctiption Factor Target data from Collection 3 in MSigDB (originally from: TRANSFAC Public, by BIOBASE, QIAGEN); version 7.4 *(BIOPAX)*
10. Human miRNA-target gene relationships from MiRTarBase; v4.5, 01-NOV-2013 (converted 13-MAR-2015) *(BIOPAX)*
11. DrugBank v4.3 converted to BioPAX from the original XML dump *(BIOPAX)*
12. Recon X: Reconstruction of the Human Genome, converted from SBML; 2.03 *(BIOPAX)*
13. Comparative Toxicogenomics Database (human), 20150603 *(BIOPAX)*
14. Small Molecule Pathway Database 2.0, 07-Jul-2015 *(BIOPAX)*
15. INOH 4.0 (signal transduction and metabolic data), 22-MAR-2011 *(BIOPAX)*
16. ChEBI Ontology v138, 01-Apr-2016 *(WAREHOUSE)*
17. UniProtKB/Swiss-Prot (human), 16-Mar-2015 *(WAREHOUSE)*
18. Selected whole-source id-mapping files (to ChEBI) from UniChem (manually edited/fixed/sorted), 29-Dec-2015 *(MAPPING)*

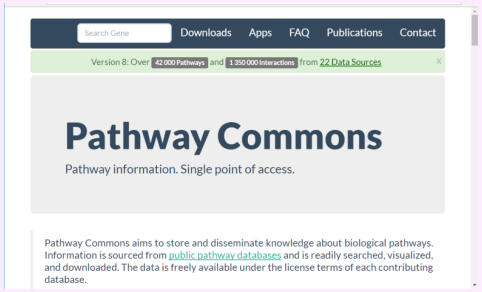
Как строить и анализировать сети взаимодействий между генами / белками ??

Интернет-доступные информационные компьютерные системы, позволяющие экстрагировать данные по связям различных типов между генами/белками

STRING - functional protein association networks
string-db.org/



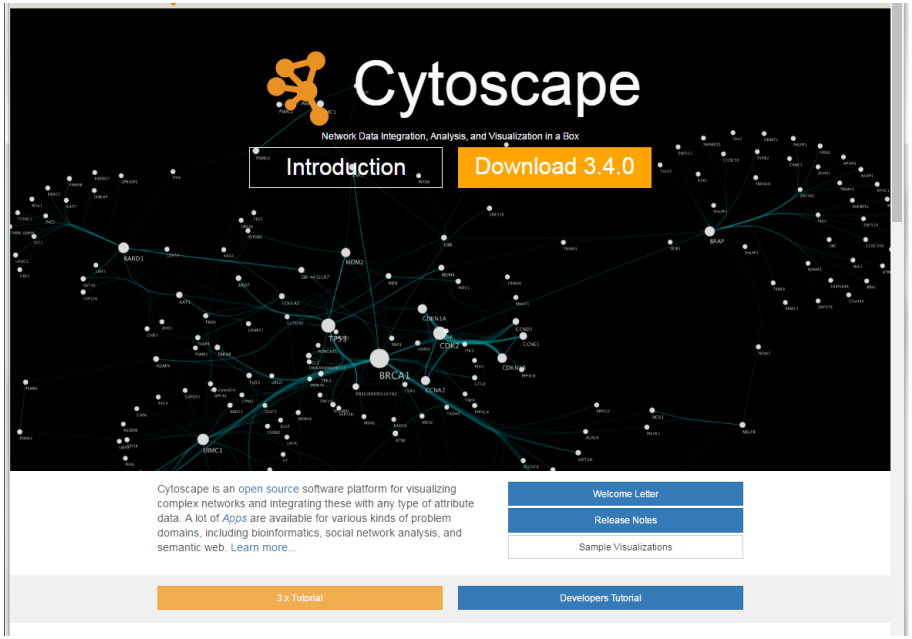
GeneMANIA
http://genemania.org/



Pathway Commons
- A Resource for Biological Pathway Analysis
http://www.pathwaycommons.org/

Cytoscape **- Network Data Integration, Analysis, and Visualization in a Box**

***Cytoscape* - *к*омпьютерная система, предназначенная для визуализации сетей межмолекулярных взаимодействий** (http://www.cytoscape.org/)



***Области использования*:**

Биология

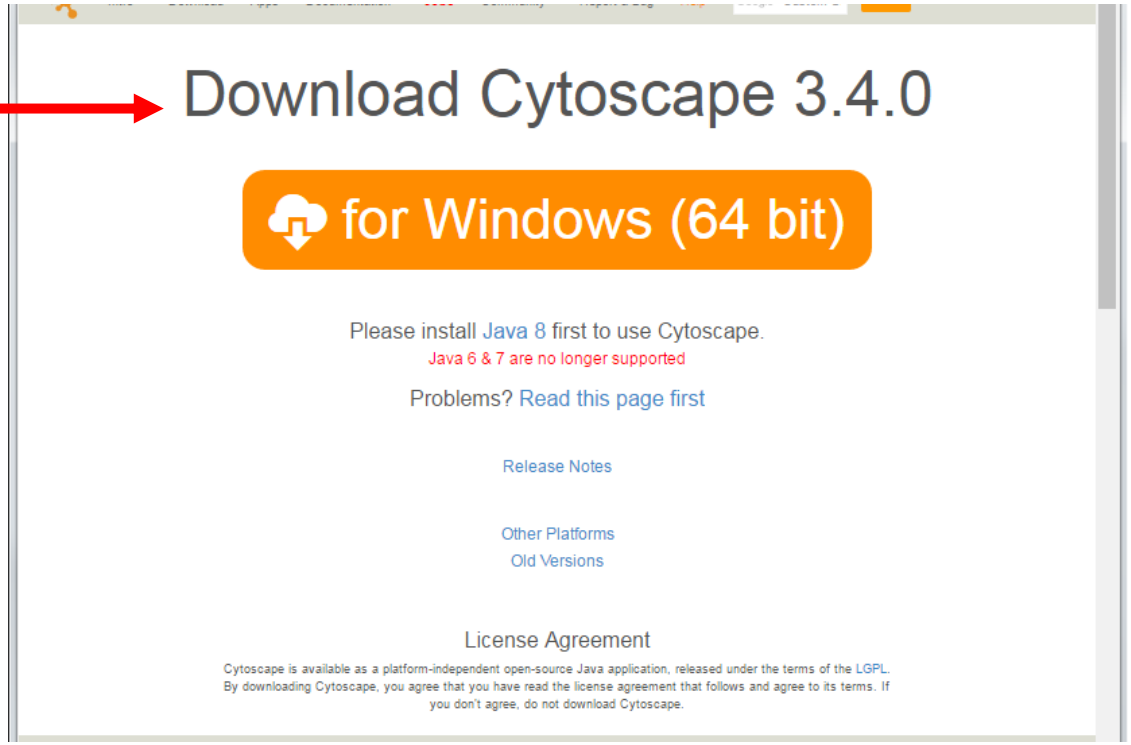
Социология (Social Science)

Общий комплексный сетевой анализ (General Complex Network Analysis)

Franz M1, Lopes CT1, Huck G1, Dong Y1, Sumer O1, Bader GD1 Cytoscape.js: a graph theory library for visualisation and analysis. Bioinformatics. 2016 Jan 15;32(2):309-11.

Lopes CT1, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD.
Cytoscape Web: an interactive web-based network browser. Bioinformatics. 2010 Sep 15;26(18):2347-8.

***Cytoscape* *бесплатная к*омпьютерная система, которую можно скачать и установить на персональном компьютере**



Основные требования к компьютеру:

- ***64-битная система***
- ***Наличие Java 8***

Franz M1, Lopes CT1, Huck G1, Dong Y1, Sumer O1, Bader GD1 Cytoscape.js: a graph theory library for visualisation and analysis. Bioinformatics. 2016 Jan 15;32(2):309-11.

Lopes CT1, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD.
Cytoscape Web: an interactive web-based network browser. Bioinformatics. 2010 Sep 15;26(18):2347-8.

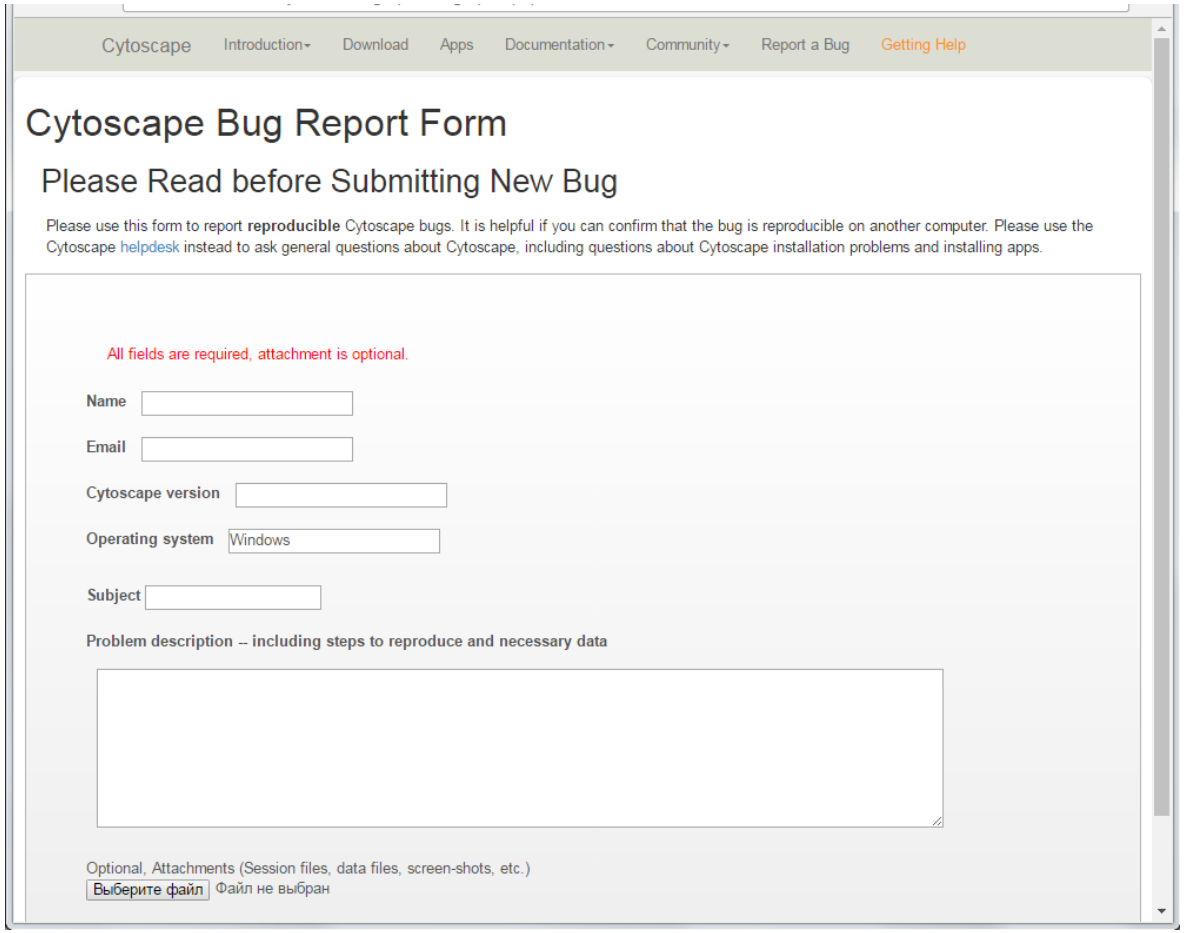
Достоинства системы Cytoscape (http://www.cytoscape.org/)

- ***Cytoscape*** **- компьютерная программа (система), (компьютерная платформа), (пакет программ) с *открытым исходным кодом,* это позволяет компьютерным специалистам всего мира участвовать в развитии *Cytoscape.***
- **Функционирует отлаженная система отслеживания ошибок.**
- **Большое количество возможностей для анализа, поскольку отлажен механизм добавления новых функциональных модулей программы в виде плагинов (апплетов = Apps (formerly called Plugins)).**
- **Каждый плагин описан в отдельной публикации в рецензируемом журнале.**
- **Одна из самых цитируемых программ (компьютерных систем)**
- **Есть возможность учитывать при визуализации дополнительные данные многих различных типов (функциональная аннотация, информация об уровне экспрессии генов и пр.)**

Franz M1, Lopes CT1, Huck G1, Dong Y1, Sumer O1, Bader GD1 Cytoscape.js: a graph theory library for visualisation and analysis. Bioinformatics. 2016 Jan 15;32(2):309-11.

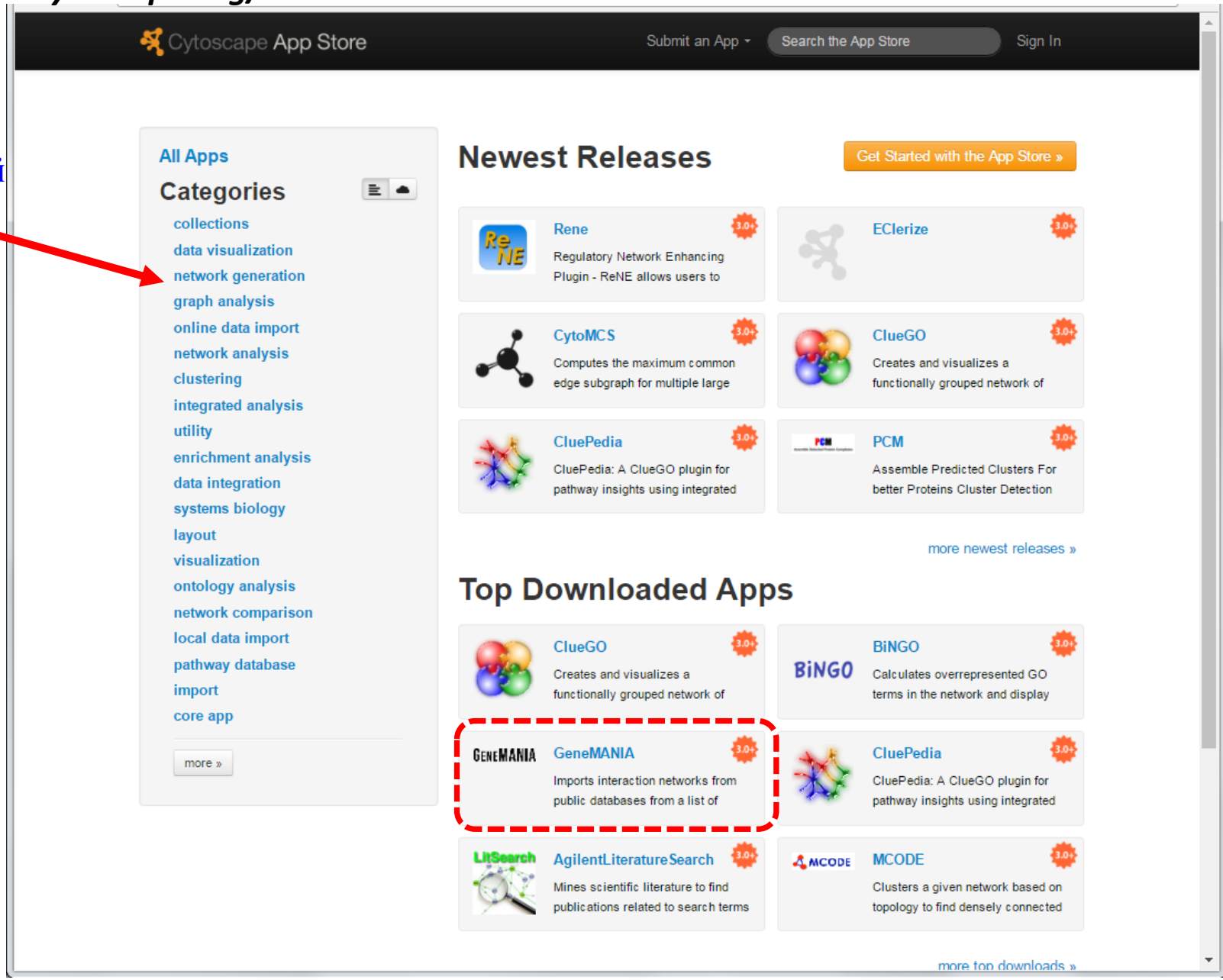
Lopes CT1, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD.
Cytoscape Web: an interactive web-based network browser. Bioinformatics. 2010 Sep 15;26(18):2347-8.

Cytoscape – функционирует отлаженная система отслеживания ошибок (багов) программы http://chianti.ucsd.edu/cyto_web/bugreport/bugreport.php

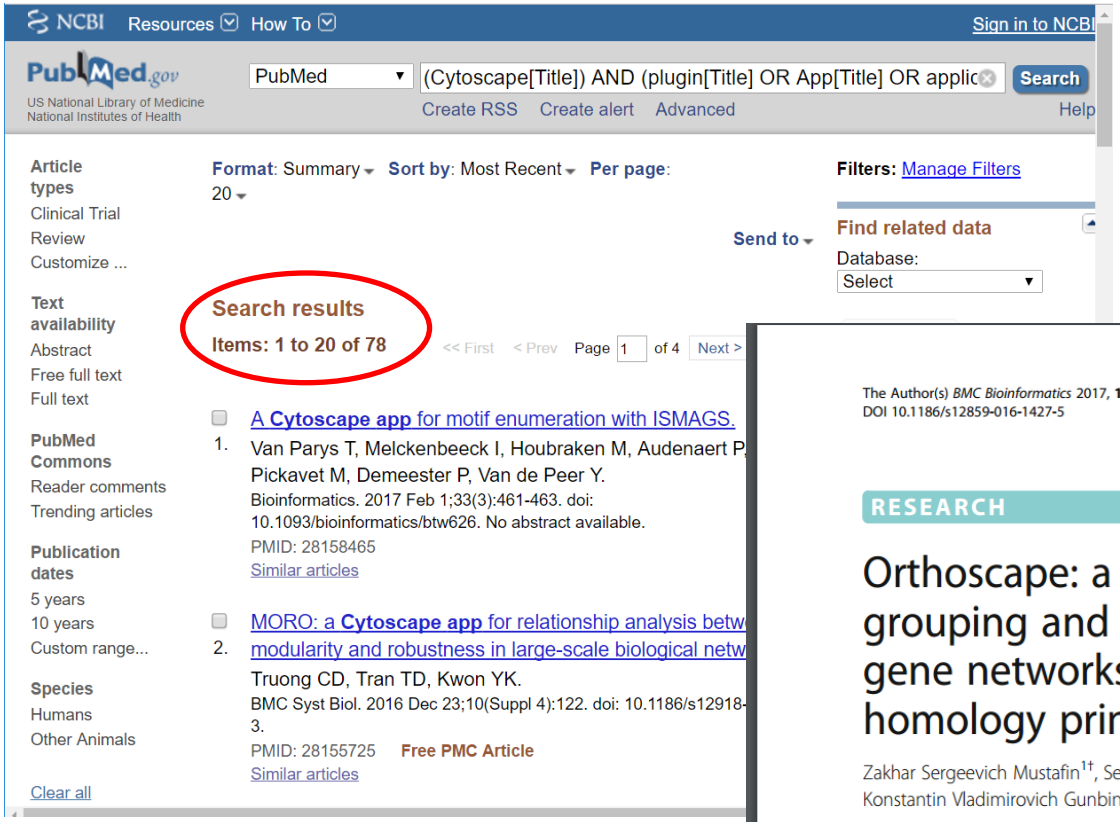


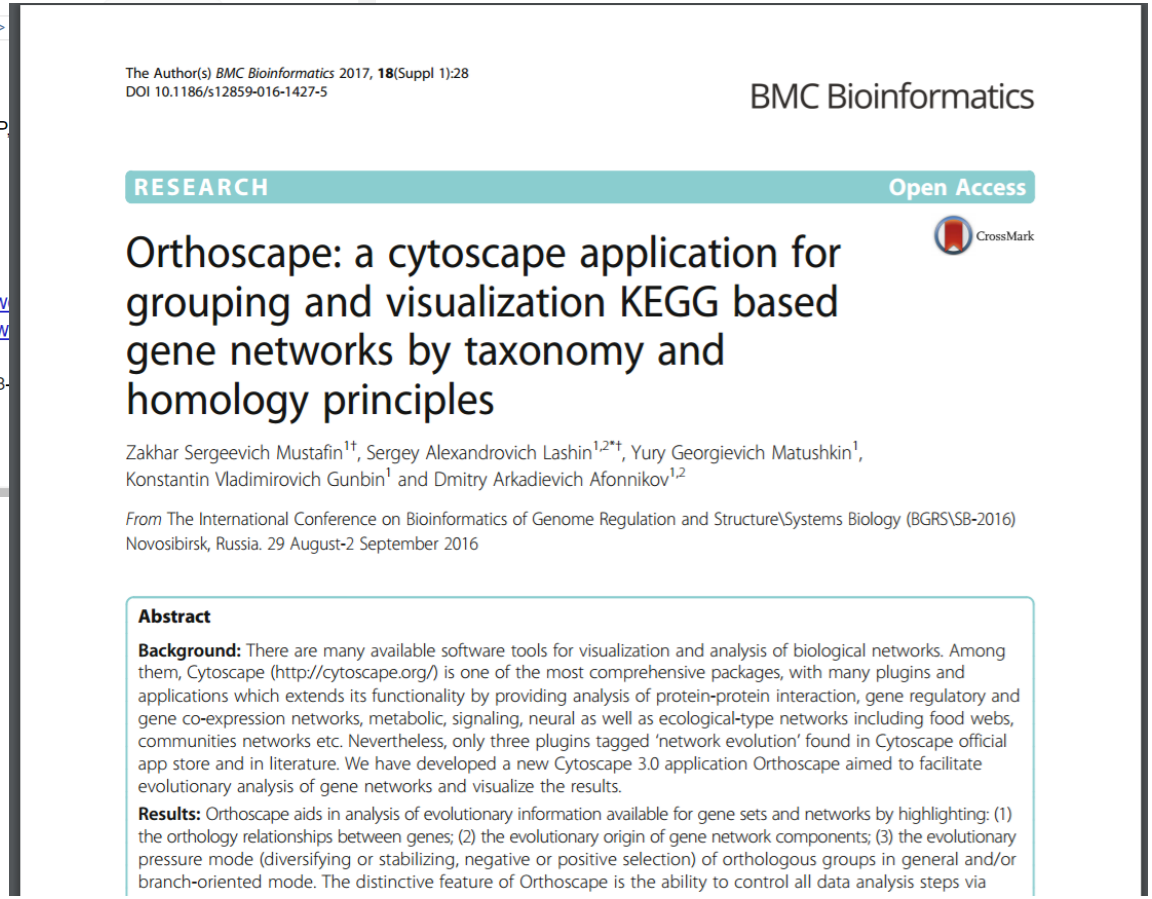
Cytoscape App Store – сайт, интегрирующий все плагины (апплеты)
http://apps.cytoscape.org/

Тематический каталог плагинов



Плагины Cytoscape получают путевку в жизнь в форме публикации в рецензируемом журнале.





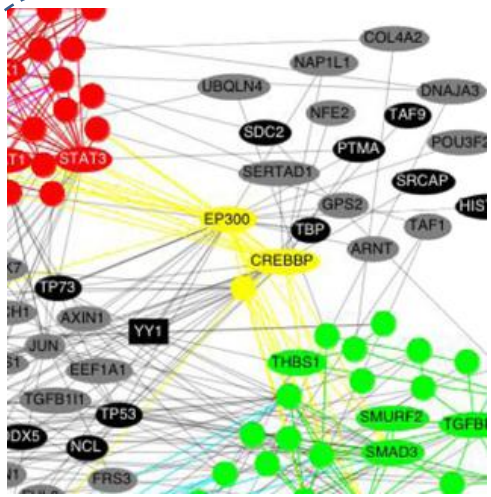
Mustafin Z.S., Lashin S.A., Matushkin Y.G., Gunbin K.V. and Afonnikov D.A. BMC Bioinformatics 2017, 18(Suppl 1):28

Возможности Cytoscape

В области молекулярной биологии, системной биологии , геномики и протеомики:
Принимать (загружать) данные о молекулярным и генетическим взаимодействиям в стандартных форматах:

Simple interaction file (SIF or .sif format)
Nested network format (NNF or .nnf format)
Graph Markup Language (GML or .gml format)
XGMML (extensible graph markup and modelling language).
SBML, BioPAX, PSI-MI Level 1 and 2.5, GraphML, Delimited text, Excel Workbook (.xls, .xlsx)

Интегрировать данные о взаимодействиях с функциональной аннотацией объектов и связей



Условные обозначения

Толщина линии соответствует уровню достоверности белок-белкового взаимодействия

- Jak/STAT
- Jak/STAT and TGFb
- TGFb
- Белок вируса HCV

Фрагмент диаграммы, отображающей белок-белковые взаимодействия между белками вируса гепатита С и белками человека

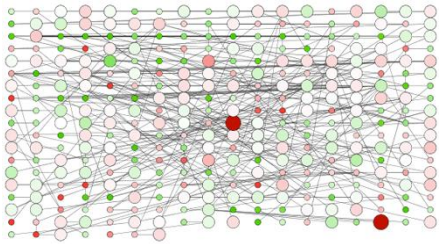
B de Chassey et al. Hepatitis C virus infection protein network.// Molecular Systems Biology (2008) 4, 230.

Возможности Cytoscape (продолжение)

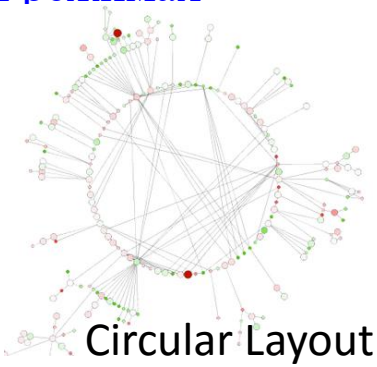
В области молекулярной биологии, системной биологии , геномики и протеомики:

Визуализация информации в различных режимах

Organic Layout



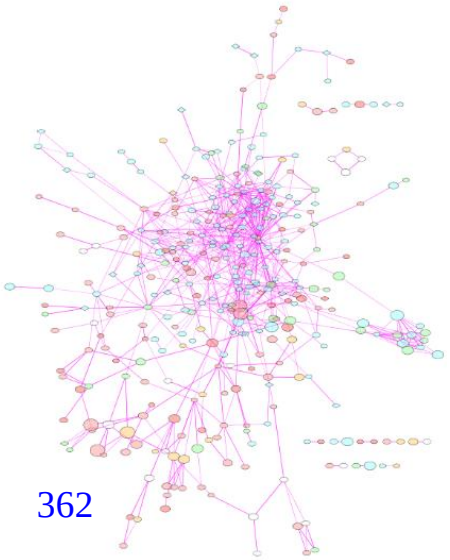
Grid Layout



Circular Layout

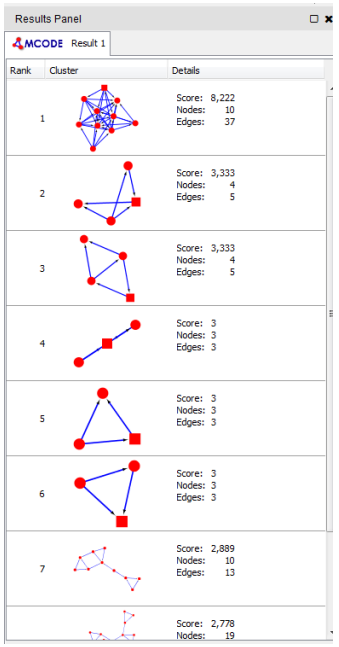
GroupAttributesLayout

Анализировать и моделировать на основе плагинов



Число вершин 362

Кластеризация с помощью программы MCODE (плагин Cytoscape)

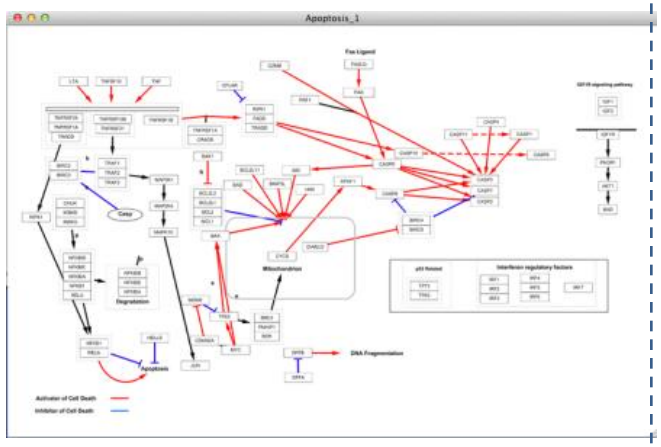


Возможности Cytoscape (продолжение)

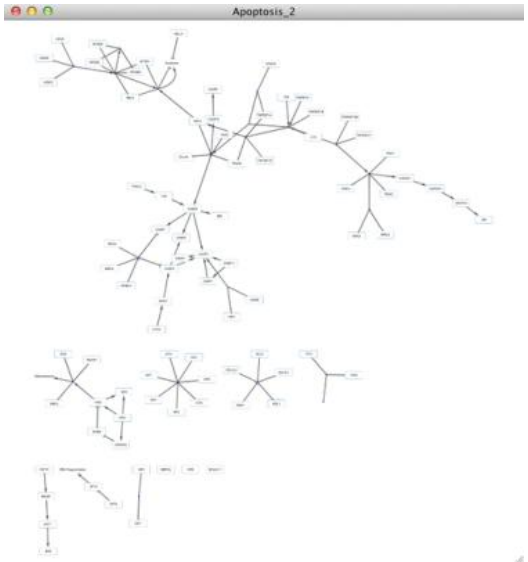
В области молекулярной биологии, системной биологии , геномики и протеомики:

Визуализировать и анализировать данные о генных сетях человека (таких как WikiPathways, Reactome, KEGG).

Диаграмма Apoptosis из WikiPathways



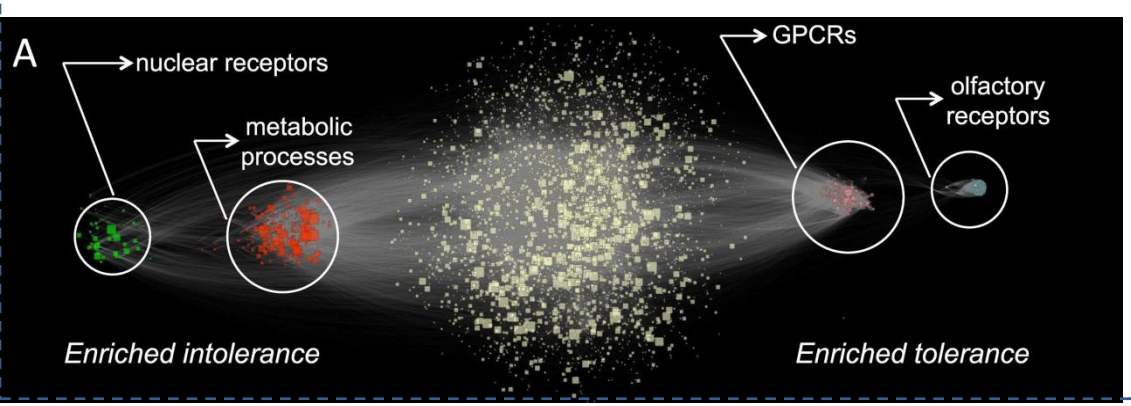
Cytoscape - режим визуализации «By default»



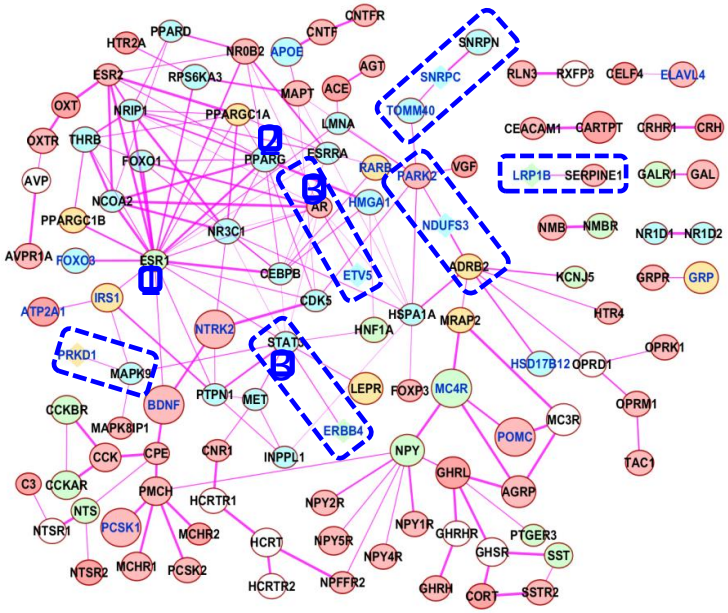
Cytoscape - режим визуализации «Import as a Network»

http://opentutorials.cgl.ucsf.edu/index.php/Tutorial:WikiPathways_App

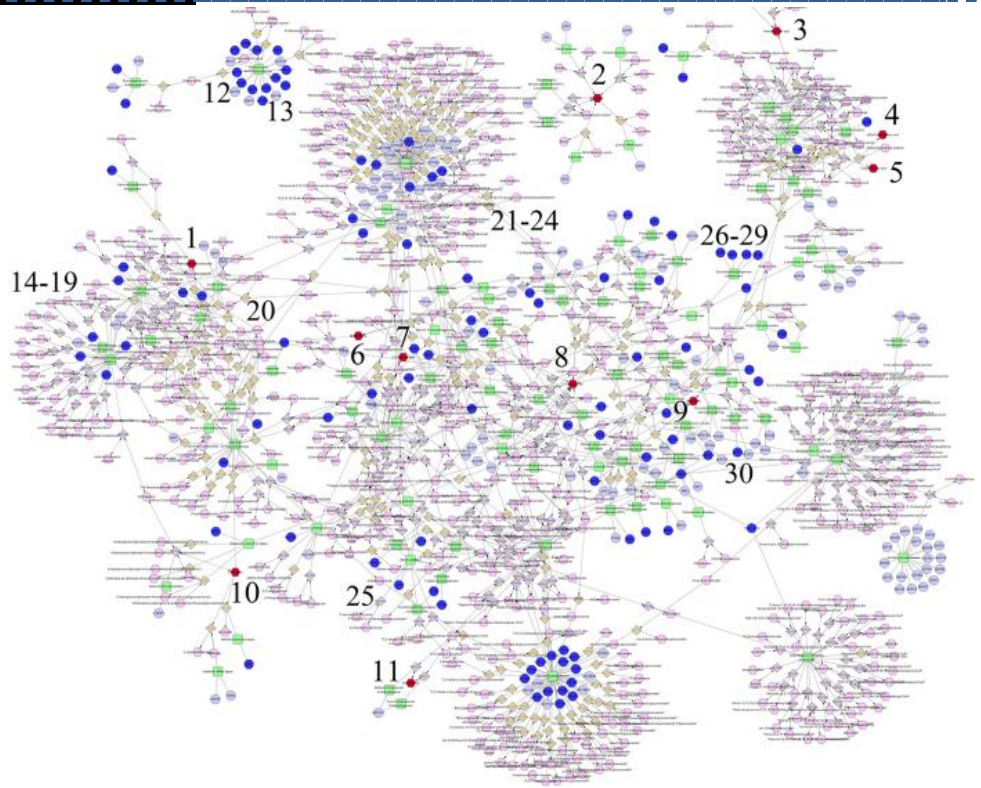
Примеры визуализации данных с помощью Cytoscape



Rackham OJ, Shihab HA, Johnson MR, Petretto E. EvoTol: a protein-sequence based evolutionary intolerance framework for disease-gene prioritization. Nucleic Acids Res. 2015;43(5):e33.



Ignatieva EV, Afonnikov DA, Saik OV, Rogaev EI, Kolchanov NA. A compendium of human genes regulating feeding behavior and body weight, its functional characterization and identification of GWAS genes involved in brain-specific PPI network. BMC Genet. 2016;17(Suppl 3):158.



A Comprehensive Analysis of Metabolomics and Transcriptomics in Cervical Cancer. Yang et al, *Sci Rep. 2017; 7: 43353*

Возможности Cytoscape (продолжение)

Общий комплексный сетевой анализ

- Может применяться в любых отраслях знаний, где используется сетевое представление данных.
- Подсчет статистики по различным характеристикам сетей
- Поиск кратчайшего пути между вершинами
- Выявление кластеров с использованием нескольких алгоритмов
- Может быть использован совместно с другими популярными программами для анализа графов (igraph, Pajek, or GraphViz)

Часть 2.

Представление генных сетей в базах данных.

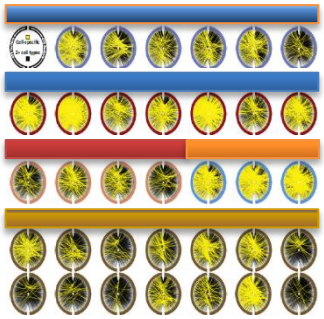
Как построить генную сеть ?

① **Эксперт-биолог строит сеть на основе анализа научных публикаций**

② **Экспериментально-теоретический подход:**
эксперимент на основе омиксных технологий
+ теоретический анализ

Эксперимент по ДНК-аза I футпринтингу
в 41 типе клеток человека
+ компьютерное предсказание
475 типов ССТФ

Сети регуляторных транскрипционных взаимодействий для 41 типов клеток человека

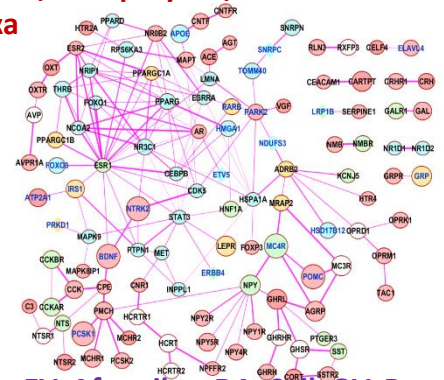


Neph S., et al., Cell, 2012, 1, 1274–1286

③ **Теоретический подход:** экстракция данных из Интернет-ресурсов, интеграция, анализ

Данные баз данных OMIM, PubMed, Human Protein Atlas, ресурсов STRING, GeneMANIA, TSEA
+ анализ системой Cytoscape

Специфичная для мозга сеть белок-белковых взаимодействий, между продуктами генов, участвующих в регуляции массы тела человека



Ignatieva EV, Afonnikov DA, Saik OV, Rogaev EI, Kolchanov NA.. BMC Genet. 2016;17(S3):158.

Информационные источники для реконструкции генной сети

Список генов, метаболических процессов, путей сигнальной трансдукции и т.п.

Информация об объектах генной сети:

- **Гены**: EntrezGene, Ensembl, GeneCards
- **Белки**: UniProtKB, PDB, NCBI Protein
- **миРНК**: miRBase, miRDB, miRTarBase
- **Метаболиты**: ChEBI, HMDB, KEGG COMPOUND, KEGG DRUG

Межмолекулярные взаимодействия

- **Тр.фактор->ген**: TRRUST, TRRD, GTRD, The Interactome
- **миРНК->ген**: miRTarBase, miRBase
- **Белок-белковые**: BioGRID, GeneMANIA, Pathway Commons, STRING

Публикации из PubMed

Данные, полученные на основе омиксных технологий

Генные сети и пути регуляции сложных биологических процессов

Функциональные модули генной сети

- **Метаболические пути**: KEGG Pathway, PANTHER Pathway, MetaCyc, BioCarta, Reactome, SMPDB, NDEx, ConsensusPathDB
- **Пути передачи сигналов**: NetPath, WikiPathways, SPIKE, SIGNOR, InnateDB
- ANDSystem
- GeneNet

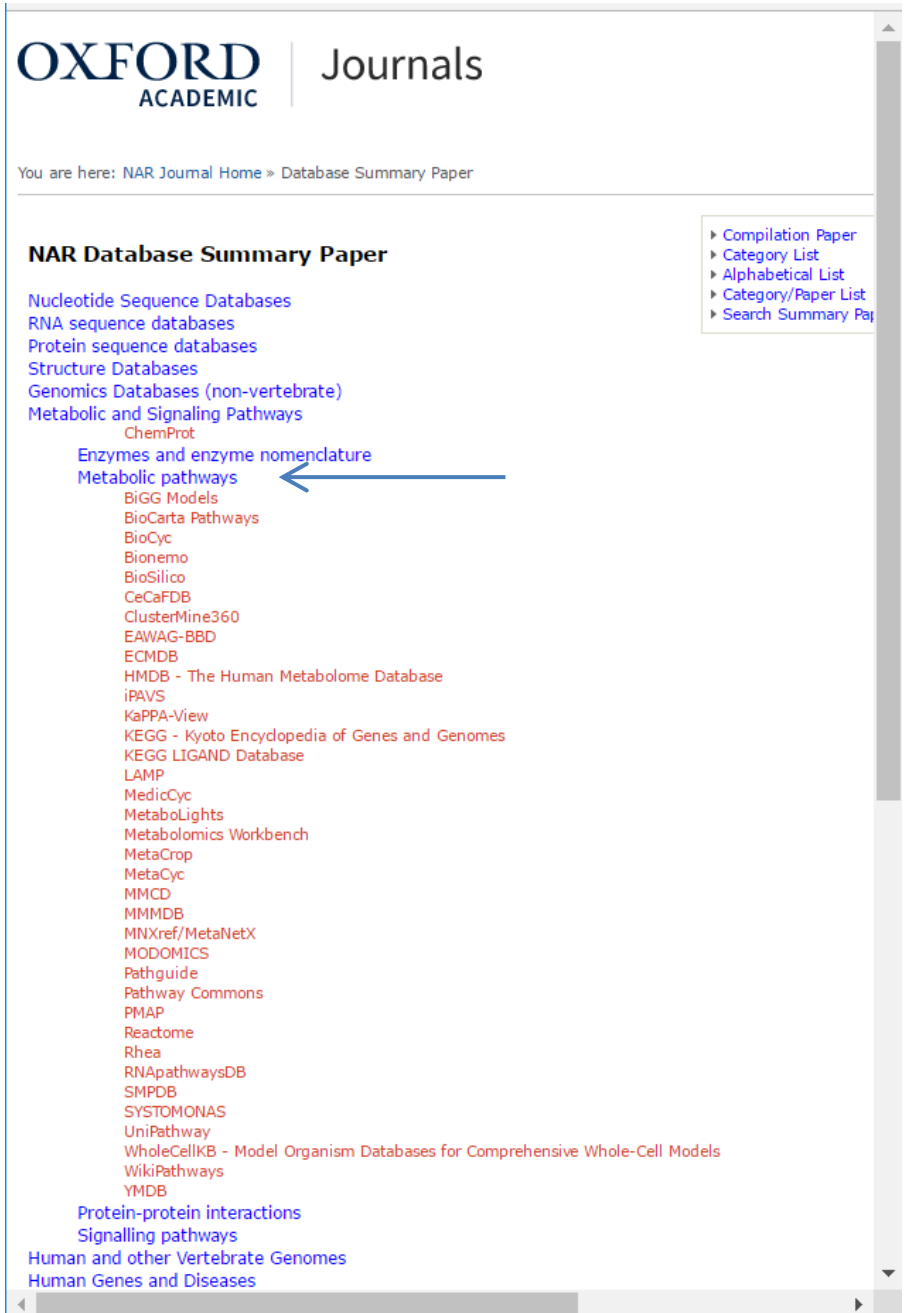
Новая генная сеть

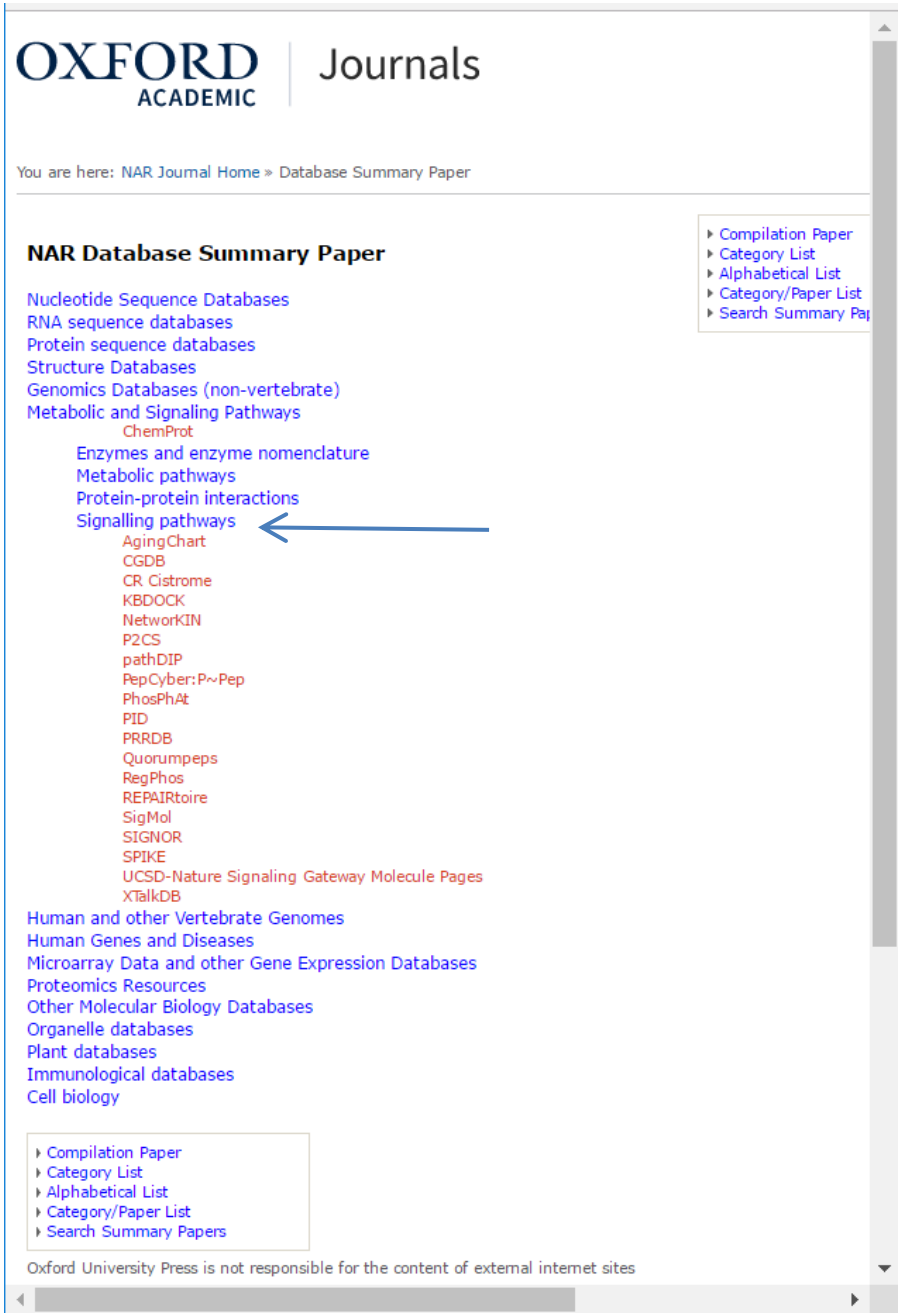
Игнатьева Е.В., Афонников Д.А., Колчанов Н.А. Интернет-доступные информационные ресурсы по генным сетям, включающие данные по человеку и животным. Вавиловский журнал генетики и селекции. 2017; 21(8):895-902
http://www.bionet.nsc.ru/vogis/download/06_Ignatjeva.pdf

Информационные ресурсы, полезные для реконструкции сетей могут содержать:

- **Характеристики объектов, являющихся компонентами генной сети (гены, белки, мРНК, миРНК, метаболиты и т.д.) (*EntrezGene, UniProtKB, miRBase, KEGG COMPOUND*)**
- **Данные о взаимодействиях между объектами в генной сети (регуляторные (ТФ->ген мишень - *The Interactome*), белок-белковые (*BioGRID*), РНК-белковые, влияние низкомолекулярных соединений на активность белков)**
- **Данные о структурных модулях генной сети (метаболические пути, сигнальные пути -*KEGG Pathway, Reactome, WikiPathways*)**
- **Диаграммы генных сетей (*WikiPathways, GeneNet*)**

Сведения о базах данных по метаболическим и сигнальным на Вэб-сайте журнала NAR (http://www.oxfordjournals.org/nar/database/c/)





Характеристика информационных ресурсов, полезных для реконструкции сетей

Игнатьева Е.В., Афонников Д.А., Колчанов Н.А. Интернет-доступные информационные ресурсы по генным сетям, включающие данные по человеку и животным. Вавиловский журнал генетики и селекции. 2017; 21(8):895-902

http://www.bionet.nsc.ru/vogis/download/06_Ignatjeva.pdf

27 информационных ресурсов.
1) типы накопленной информации;
2) способы представления информации;
3) способы наполнения баз данных;
4) основные источники информации;
5) про-граммные средства, позволяющие осуществлять поиск и анализ данных

Информационное содержание баз в 2017 г. описано в ПРИЛОЖЕНИИ

http://www.bionet.nsc.ru/vogis/download/pict-2017-21/appx13.pdf

Таблица 1. Интернет-доступные базы данных по генным сетям, метаболическим и сигнальным путям, путям регуляции других биологических процессов и молекулярно-генетическим взаимодействиям (представлены в алфавитном порядке)

Название базы	Способ наполнения или источник данных	Тип данных	Ссылка
ANDSystem	БД, ААТ	ББВ, РВ, коэкспрессия, участие белков в биологических процессах по данным UniProt-GOA, использование вещества для лечения заболевания, ассоциации «ген–заболевание»	http://www-bionet.sscc.ru/andvisio/
BiGG Models	Р	МП	http://bigg.ucsd.edu/
BioCarta	Р	МП, ППС, ПРБП, заболевания	https://cgap.nci.nih.gov/Pathways/BioCarta_Pathways
BioCyc	Р, БД, КП	МП	https://biocyc.org/
BioGRID	Р	ББВ, ГВ	https://thebiogrid.org/
ConsensusPathDB	Р, БД	МП, ППС, ПРБП, ББВ, РВ, ГВ	http://cpdb.molgen.mpg.de/
GeneMANIA	ААТ, БД, КП	ББВ, ГВ, коэкспрессия, связи между генами/белками, предсказанные компьютерными методами	http://genemania.org/
GeneNet	Р	Структурно-функциональная организация ГС, ППС, МП	http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/
GTRD	КАОД	ДНК-белковые взаимодействия	http://gtrd.biouml.org
InnateDB	Р, БД	ППС	http://www.innatedb.ca/
The Interactome	КАОД	Тканеспецифические СТР	http://www.regulatorynetworks.org/
KEGG PATHWAY	Р	МП, ППС, ПРБП, заболевания, лекарства	http://www.genome.ad.jp/kegg
miRBase	КАОД, КП, Р	РВ	http://www.mirbase.org/
MetaCyc	Р	МП	http://metacyc.org/
NDEx	БД, Р	МП, ППС, ПРБП, ББВ, РВ	http://www.ndexbio.org/#/
NetPath	Р	ППС человека	http://www.netpath.org/
PANTHER Pathway	Р, БД	МП, ППС	http://pantherdb.org/
Pathway Commons	БД	ББВ, РВ	http://www.pathwaycommons.org/
Reactome	Р	МП, ППС, пути транспорта молекул в клетке и репликации ДНК	http://www.reactome.org/
SIGNOR	Р	ППС и регуляторные взаимодействия между их участниками, ПРБП, заболевания	http://signor.uniroma2.it/
SMPDB	Р	МП, ППС, ПРБП, заболевания	http://smpdb.ca/
SPIKE	Р, БД	ППС человека	http://www.cs.tau.ac.il/~spike/
STRING	Р, ААТ, БД, КП	ББВ, коэкспрессия, связи между генами/белками, предсказанные компьютерными методами	string-db.org/
TRED	Р, БД, КП	СТР, РВ	http://rulai.cshl.edu/TRED
TRRD	Р	РВ, ДНК-белковые взаимодействия	http://wwwmgs.bionet.nsc.ru/mgs/gnw/trrd/
TRRUST	ААТ, Р	СТР, РВ	http://www.grnpedia.org/trrust
WikiPathways	Р	МП, ППС, ПРБП, заболевания	https://www.wikipathways.org/index.php/WikiPathways

Примечание. Способы наполнения: Р – ручное аннотирование научных публикаций; ААТ – автоматический анализ текстов; БД – импорт из других баз данных; КП – компьютерные предсказания; КАОД – компьютерный анализ данных, полученных высокопроизводительными экспериментальными методами. Типы данных: ББВ – белок-белковые взаимодействия; ГВ – генетические взаимодействия; МП – метаболические пути; ППС – пути передачи сигналов; ПРБП – пути регуляции биологических процессов на клеточном и организменном уровне; РВ – регуляторные взаимодействия (транскрипционный фактор – регулируемый ген либо миРНК – регулируемый ген); СТР – сети транскрипционной регуляции. Более подробное описание баз данных со ссылками на наиболее свежие публикации представлено в Приложении 2.

В лекции № 4 будет дана характеристика баз данных по генным сетям, а также по метаболическим и сигнальным путям:

1) GeneNet – ИЦиГ СО РАН , г.Новосибирск
http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/

2) KEGG Kyoto encyclopedia of genes and genomes:
integrated suite of databases on genes, proteins, and metabolic pathways
http://www.genome.ad.jp/kegg

3) MetaCyc Metabolic Database http://metacyc.org/
+ BioCyc (Database Collection) https://biocyc.org/

4) Reactome http://www.reactome.org/

5) WikiPathways http://www.wikipathways.org/index.php/WikiPathways

7) Signor http://signor.uniroma2.it/

8) SPIKE http://www.cs.tau.ac.il/~spike/

8) BioCarta https://cgap.nci.nih.gov/Pathways/BioCarta_Pathways

GeneNet:
публикация в Nucleic Acids Research 2005

Nucleic Acids Research, 2005, Vol. 33, Database issue ***D425–D427***
doi:10.1093/nar/gki077

GeneNet in 2005

E. A. Ananko*, N. L. Podkolodny, I. L. Stepanenko, O. A. Podkolodnaya, D. A. Rasskazov, D. S. Miginsky, V. A. Likhoshvai, A. V. Ratushny, N. N. Podkolodnaya and N. A. Kolchanov

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Lavrentiev Avenue 10, Novosibirsk 630090, Russia

Received September 15, 2004; Revised and Accepted October 8, 2004

ABSTRACT

The GeneNet system is designed for collection and analysis of the data on gene and metabolic networks,

of the manifold data on the expressio
and changes in the concentration of
macromolecular interactions, enzymatic


Nucleic Acids Research, 2005, Vol. 33, Database issue ***D425–D427***
doi:10.1093/nar/gki077

GeneNet in 2005

E. A. Ananko*, N. L. Podkolodny, I. L. Stepanenko, O. A. Podkolodnaya, D. A. Rasskazov, D. S. Miginsky, V. A. Likhoshvai, A. V. Ratushny, N. N. Podkolodnaya and N. A. Kolchanov

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Lavrentiev Avenue 10, Novosibirsk 630090, Russia

Received September 15, 2004; Revised and Accepted October 8, 2004

ABSTRACT

The GeneNet system is designed for collection and analysis of the data on gene and metabolic networks, signal transduction pathways and kinetic characteristics of elementary processes. In the past 2 years, the GeneNet structure was considerably improved: (i) the current version of the database is now implemented using ORACLE9i; (ii) the capacities to describe the structure of the protein complexes and the interactions between the units are increased; (iii) two tables with kinetic constants and more detailed descriptions of certain reactions were added; and (iv) a module for kinetic modeling was supplemented. The current SRS release of the GeneNet database contains 37 graphical maps of gene networks, as well as descriptions of 1766 proteins, 1006 genes, 241 small molecules and 3254 relationships between gene network units, and 552 kinetic constants. Information distributed between 16 interlinked tables was obtained by annotating 1980 journal publications. SRS release of the GeneNet database, the graphical viewer and the modeling section are available at http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/.

INTRODUCTION

Systematic arrangement and analysis of a variety of data subsets present a challenge. The GeneNet system (1) was developed to respond to the challenge. The widely available specialized databases, such as KEGG (2), BioCarta (http://www.biocarta.com/index.asp), BIND (3), TRANSPATH (4) and MetaCyc (5), among others, are concerned with metabolic or signal transduction pathways. The GeneNet workbench is more versatile, enabling the description of functions and regulation of complicated biological systems on the basis of the manifold data on the expression regulation of genes and changes in the concentration of their products, the macromolecular interactions, enzymatic reactions, the effects of external agents, signal transduction pathways, to name a few. The GeneNet diagrams represent mainly the structural–functional organization of the gene network (1) that controls particular processes in eukaryotes (6). The GeneNet's task is not to give a detailed description of protein–protein interactions or of the gene or protein structure; all this is available from the other established databases. The intention is to describe, in more detail, the relationships between the gene network units and regulatory influences on the relationships (7). The information from the GeneNet database is further used in the developing of kinetic computer models of various biological processes (8).

GeneNet MODULES

The GeneNet system consists of the following functional modules:

(i) a database that compiles information on the structural and functional organization of the gene and metabolic networks, their elementary units (proteins, genes, RNAs, small molecules, etc.) and elementary interactions between the units (1);
(ii) a graphical viewer that allows display of the structure of the gene network and interactions of its units as a two-dimensional graph;
(iii) a section of computer modeling of gene networks;
(iv) software for analysis of the graph structure of gene network and its functional characteristics (9);
(v) a graphical editor enabling the construction of new diagrams and to modify the existing ones, to add new information to the database.

It should be noted that the user can request the software for analysis and the graphical editor in the GeneNet licensed version only. For this reason, the descriptions of these tools are omitted here. Furthermore, the licensed GeneNet version

*To whom correspondence should be addressed. Tel: +7 3832 333119; Fax: +7 3832 331278; Email: ananko@bionet.nsc.ru

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use permissions, please contact journals.permissions@oupjournals.org.

© 2005, the authors

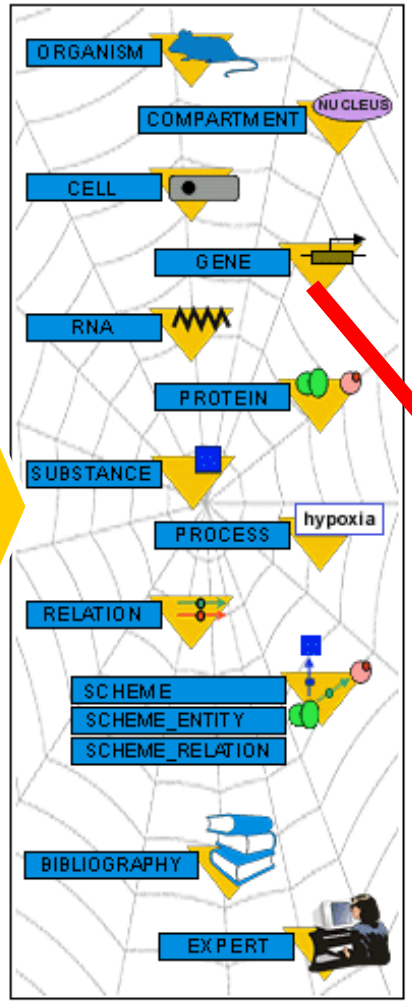
Nucleic Acids Research, Vol. 33, Database issue © Oxford University Press 2005; all rights reserved

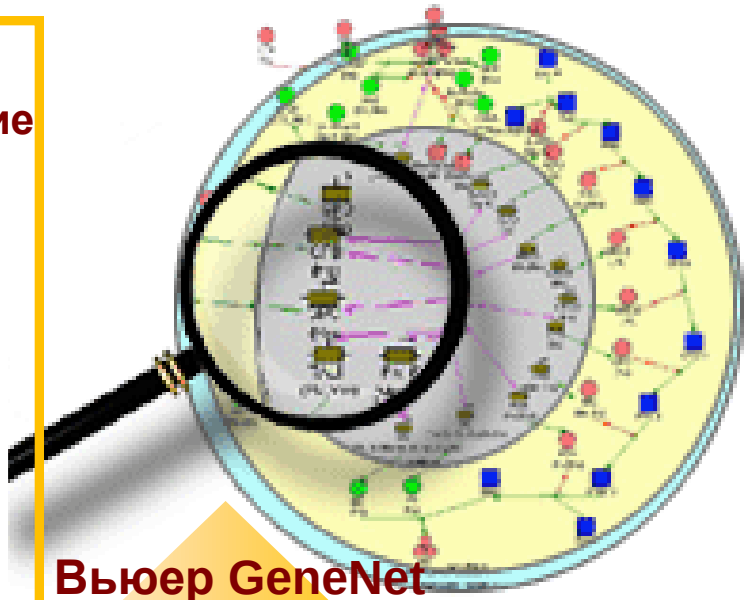
Технология GeneNet: основные модули, которыми пользуются разработчики базы

Статьи из журналов с описанием экспериментов

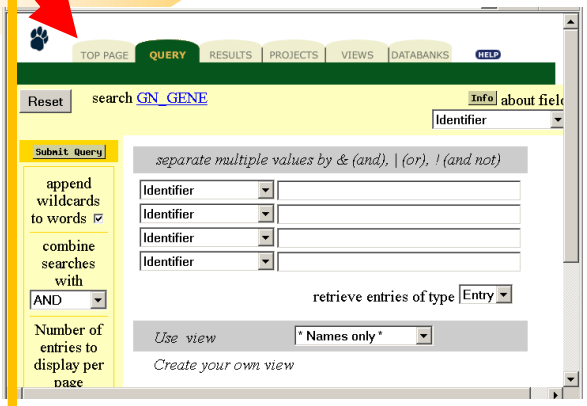
Графический редактор GeneEd для внесения новой информации

База данных (текстовое описание объектов, связей, диаграмм)



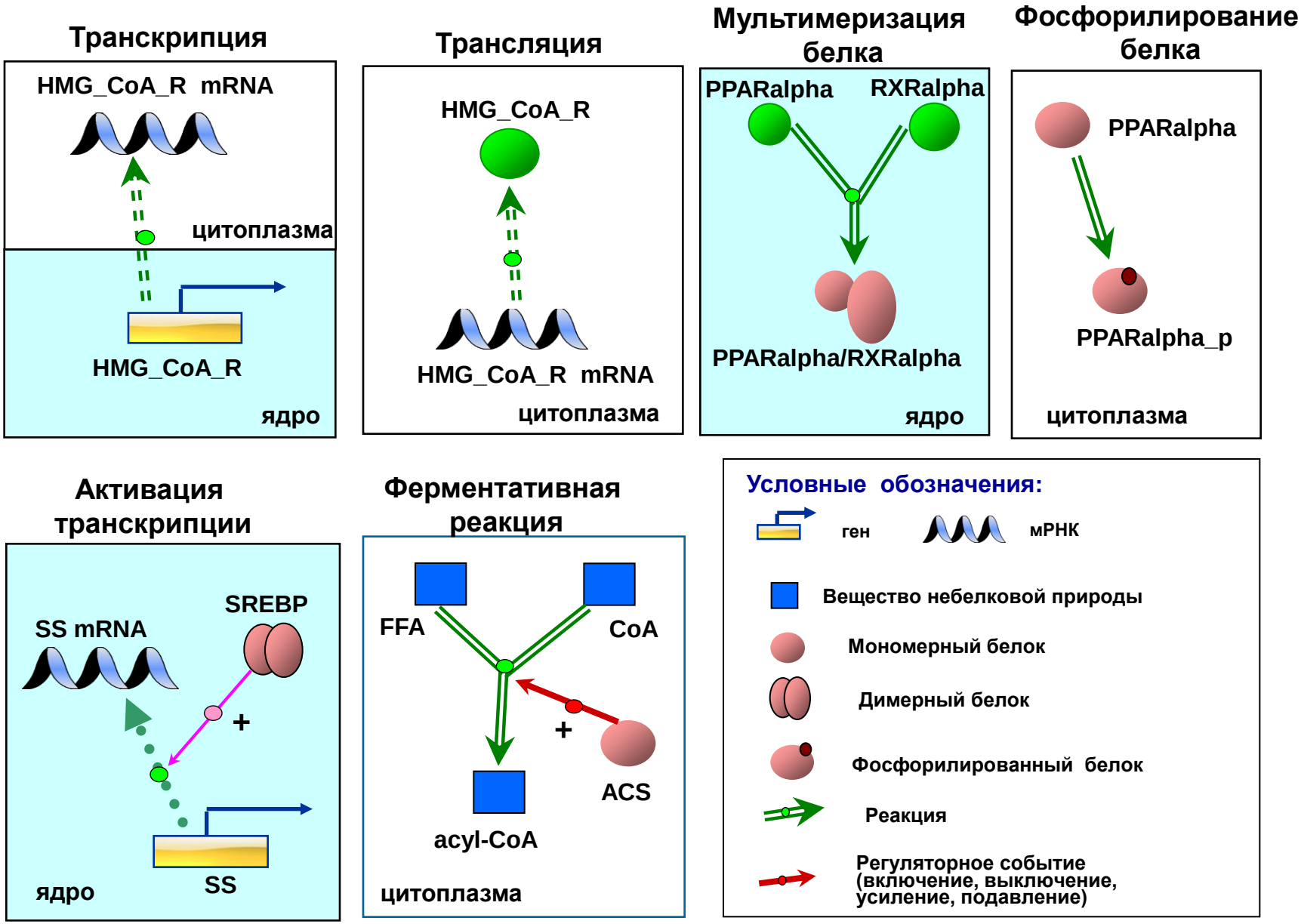


Вьюер GeneNet
= редактор GeneEd, у которого отключена опция редактирования

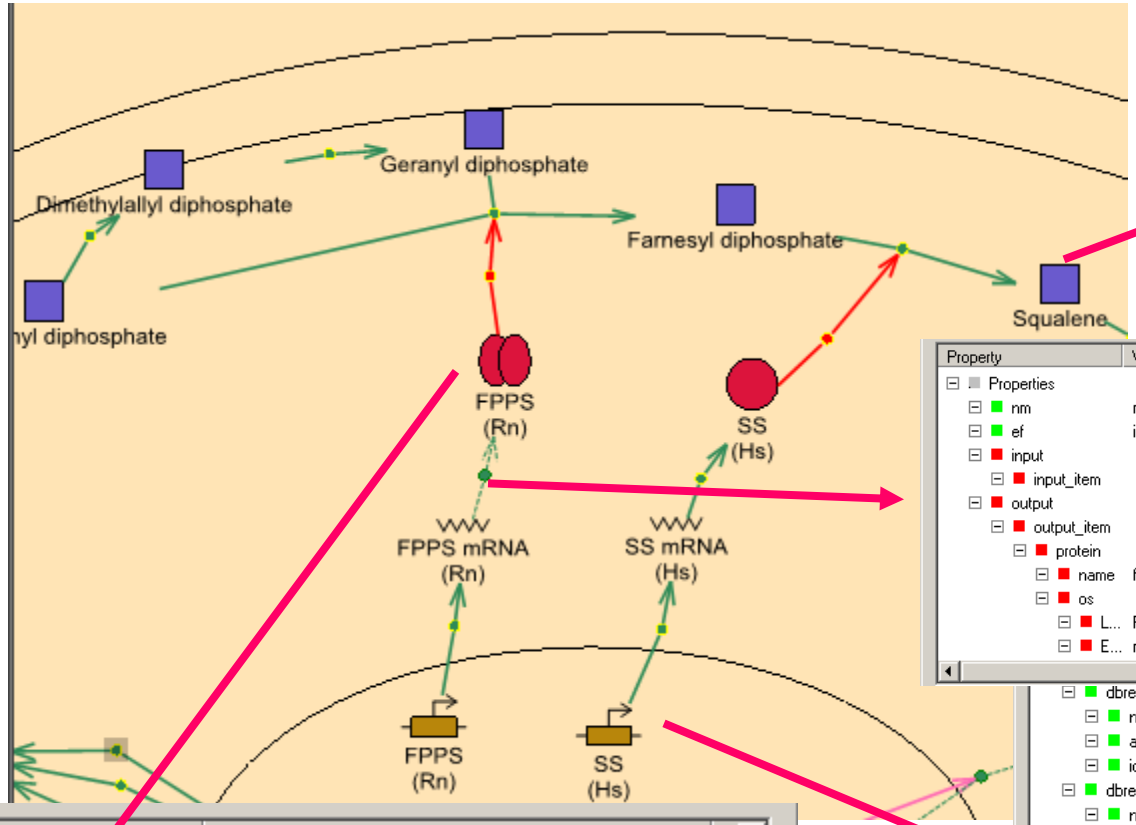


Доступ к информации через поисковую систему SRS

GeneNet: примеры графического изображения элементарных процессов (повторение)



БАЗА GeneNet (ИЦиГ СО РАН): представление данных в графическом редакторе



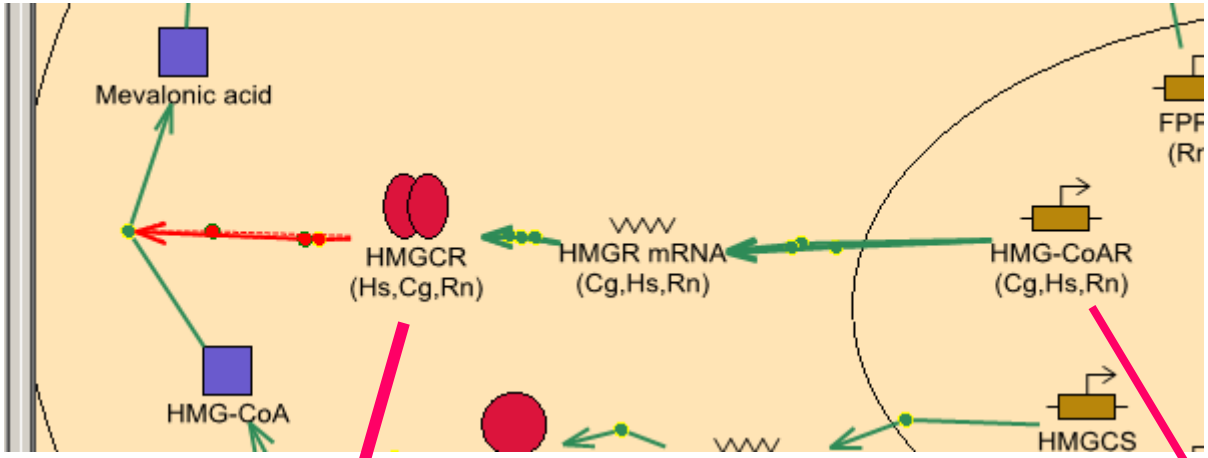
Property	Value
Properties	
sn	Squalene
nm	Squalene

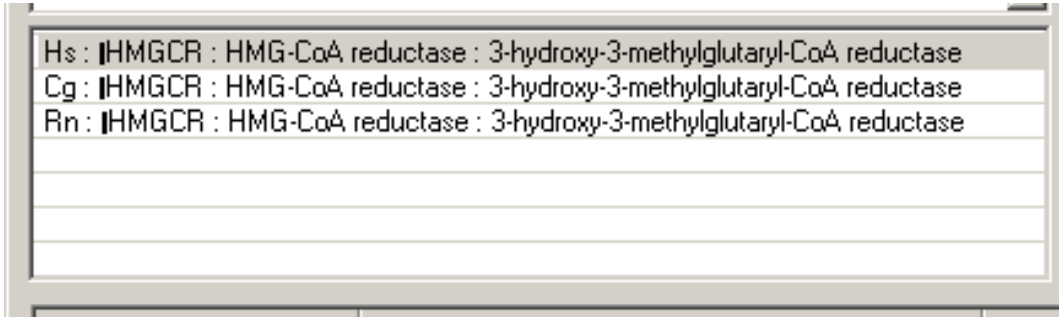
Property	Value
Properties	
nm	reaction
ef	indirect
input	
input_item	
output	
output_item	
protein	
name	farnesyl diphosphate synthase
os	
L...	Rattus norvegicus
E...	rat

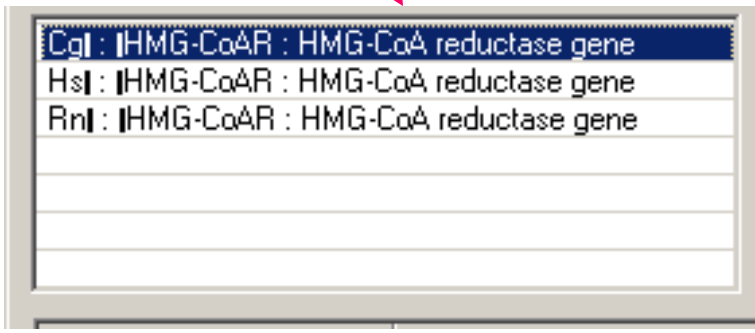
Property	Value	Comment
dbref		
name	EMBL	Database name
ac	U18994	accession number of the ...
id	HS189941	identifier of the entry in th...
dbref		DataBase Reference
name	SWISS-PROT	Database name
ac	P37268	accession number of the ...
id	FDFT_HUMAN	identifier of the entry in th...
inducer		Inducer
name	cholesterol (repression)	full name
cell		Cell
sn	HepG2	short name
de	HepG2 hepatoma cell line	description
os		Organism Species
Latin	Homo sapiens	Species Latin name
English	human	Species English name
cell		Cell
sn	liver	short name
de	liver	description

Property	Value
Properties	
sn	FPPS
nm	farnesyl diphosphate synthase
sy	geranyl-diphosphate: isopentenyl-diphosphate ger...
os	
Latin	Rattus norvegicus
English	rat
dbref	

БАЗА GeneNet (ИЦиГ СО РАН)
представление данных об объектах разных видов в графическом редакторе



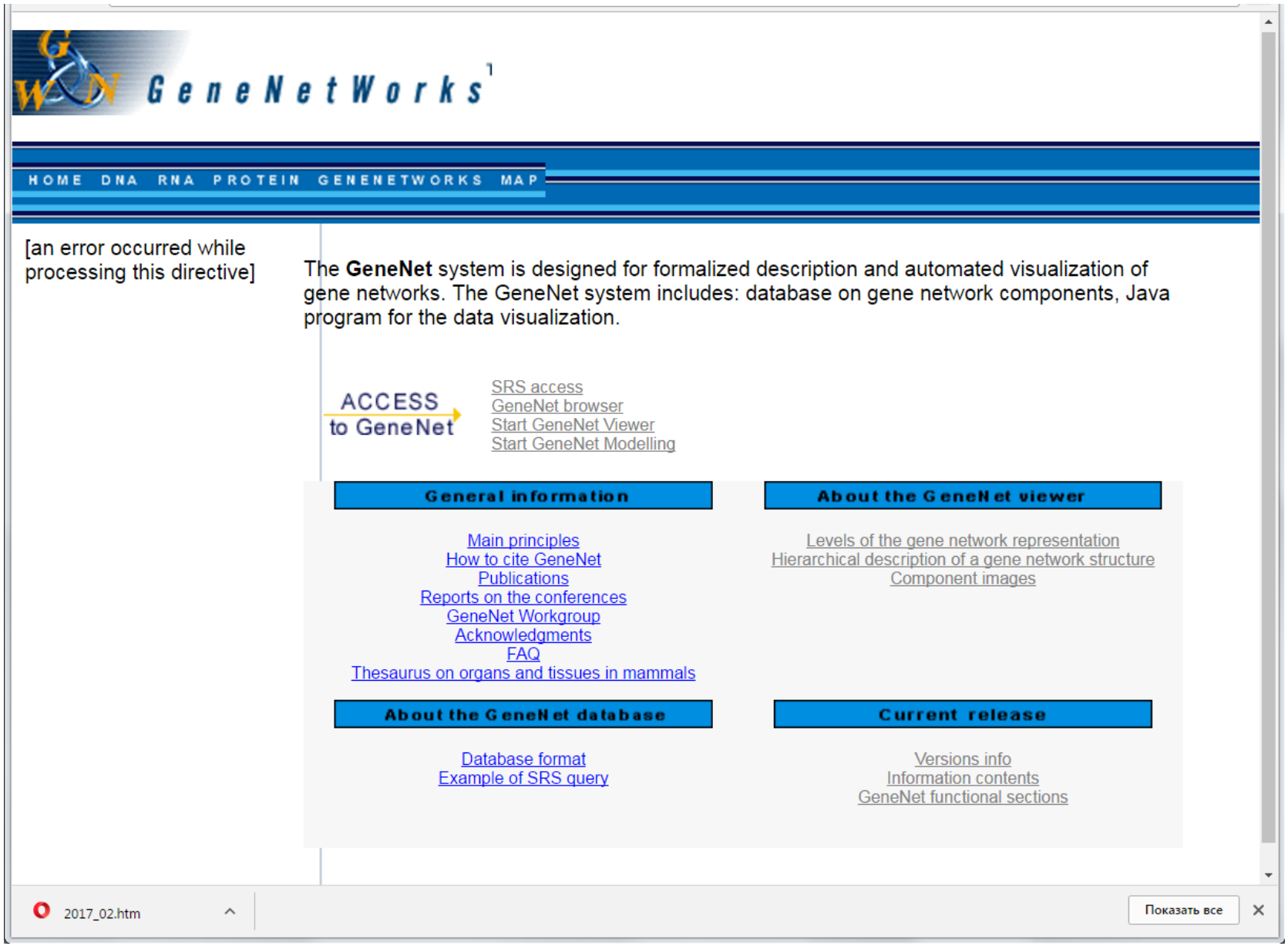




Группа одноименных (гомологичных) объектов разных видов (гены, мРНК, белки) представлены на диаграмме в базе GeneNet одним образом. Имеется возможность просмотреть текстовую информацию о каждом объекте

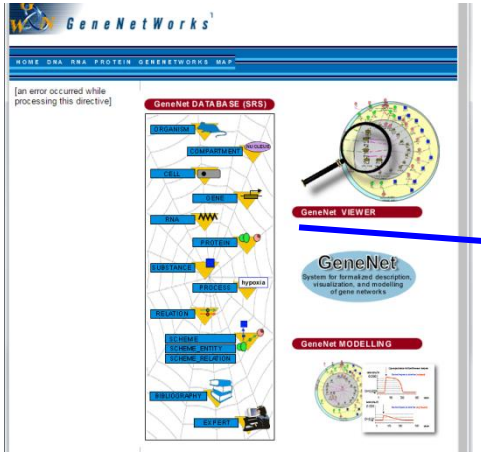
Интернет-доступная версия GeneNet

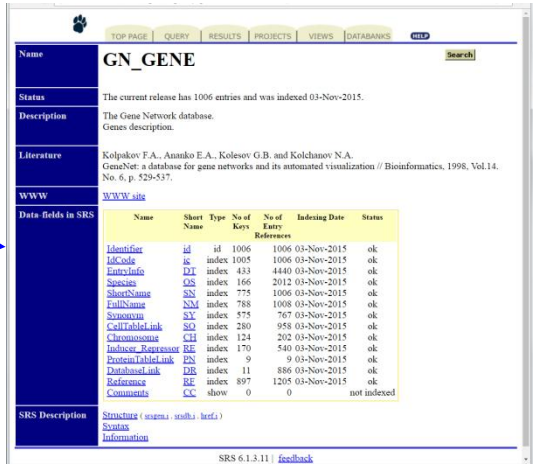
http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/



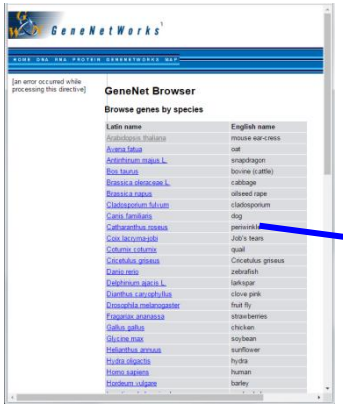
Интернет-доступная версия GeneNet: основные модули

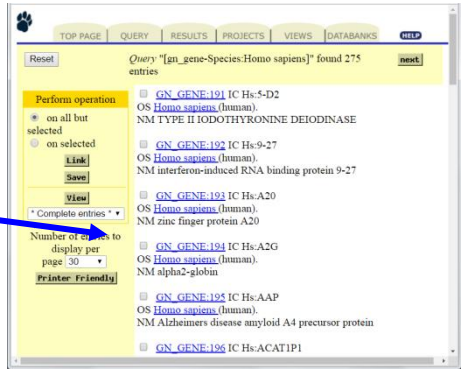
База GeneNet, доступная для поиска через систему SRS



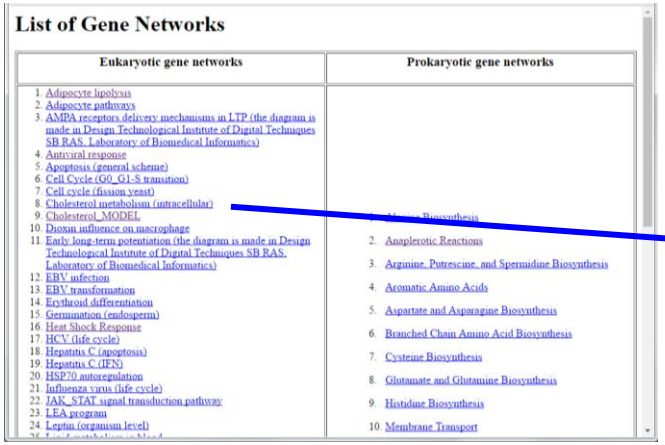


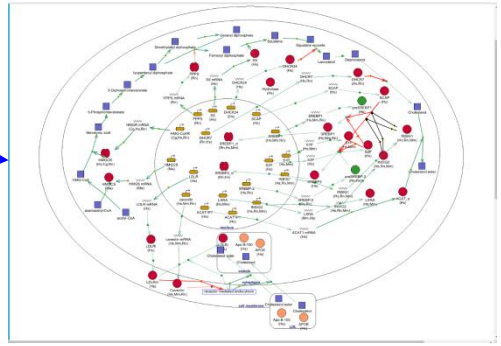
Браузер базы GeneNet – список видов организмов и соответствующих им генов





Список диаграмм





GeneNet: диаграмма, описывающая противовирусный ответ

GeneNetWorks

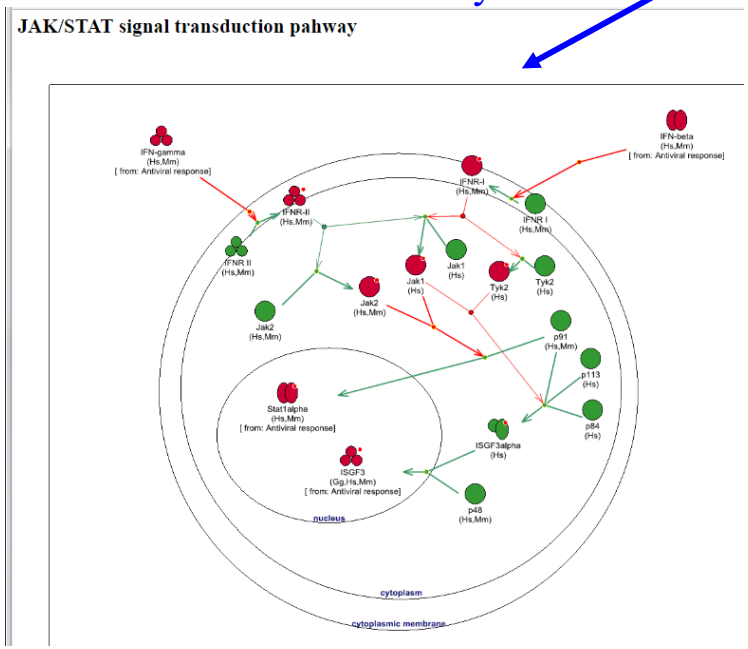
Antiviral response

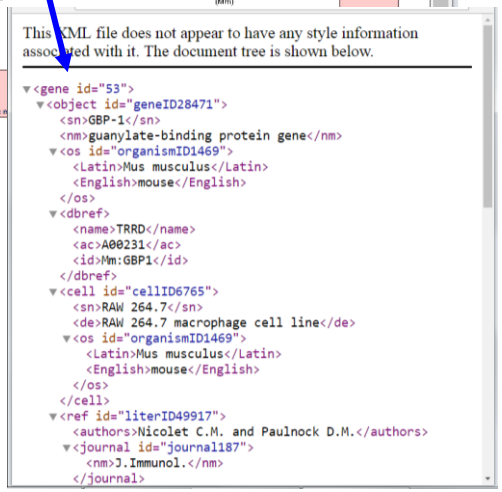
INF-gamma

INF-beta

Гиперссылка на подсхему

JAK/STAT сигнальный путь





Информационное содержание интернет-доступной версии GeneNet

GeneNetWorks

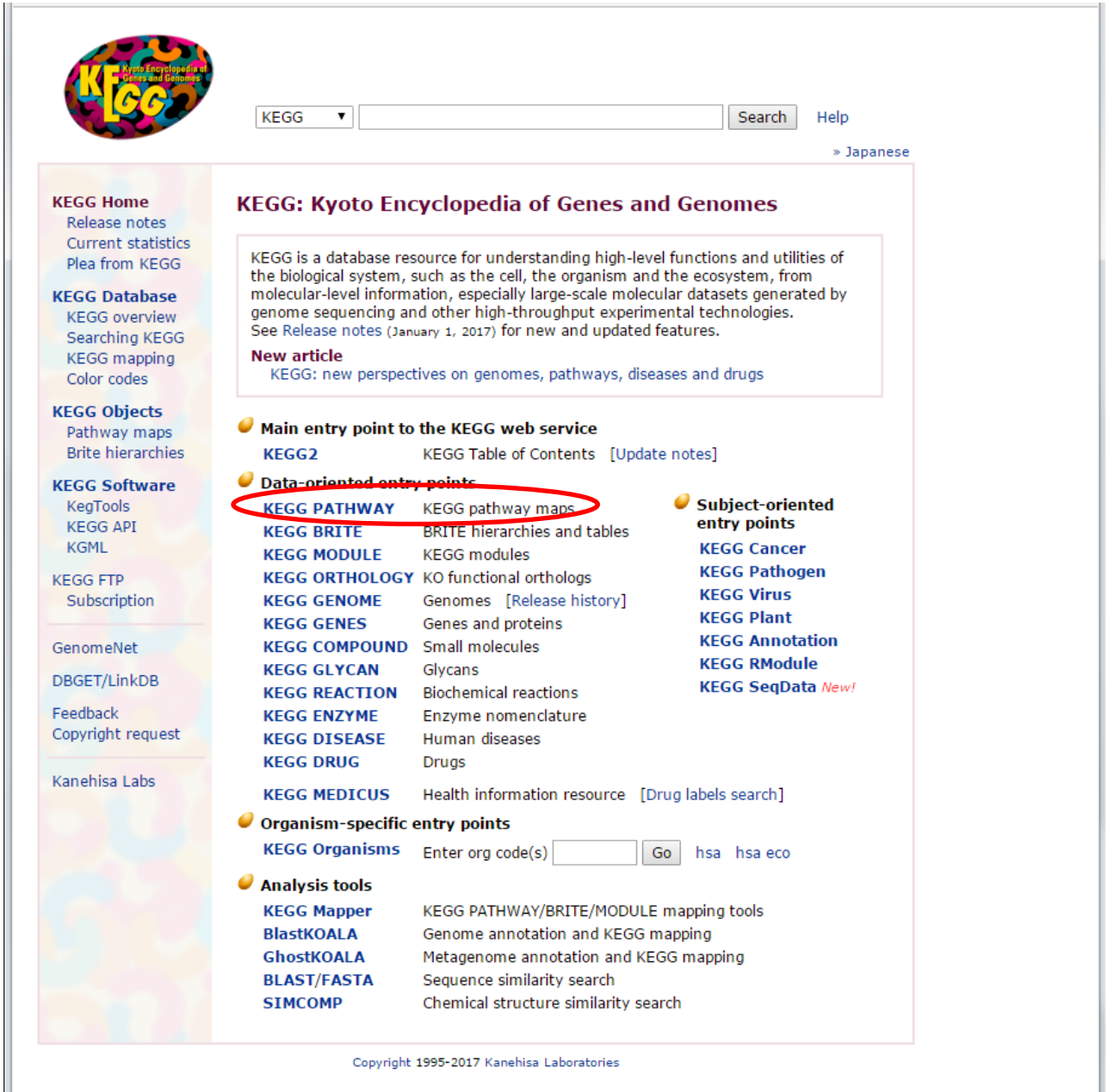
39 диаграмм
1006 генов
1766 белков
3634 связей
93 вида организмов

Тематические секции

Cell cycle
Lipid metabolism
Endocrine regulation
Erythrocyte maturation
Immune system
Plant genes networks
Stress response
Redox-regulation

KEGG: Kyoto Encyclopedia of Genes and Genomes

http://www.genome.jp/kegg/



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

Select prefix: map | Organism | Enter keywords | Go | Help

[New pathway maps | Update history]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks for:

1. **Metabolism**
   Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan
   Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
2. **Genetic Information Processing**
3. **Environmental Information Processing**
4. **Cellular Processes**
5. **Organismal Systems**
6. **Human Diseases**

and also on the structure relationships (KEGG drug structure maps) in:

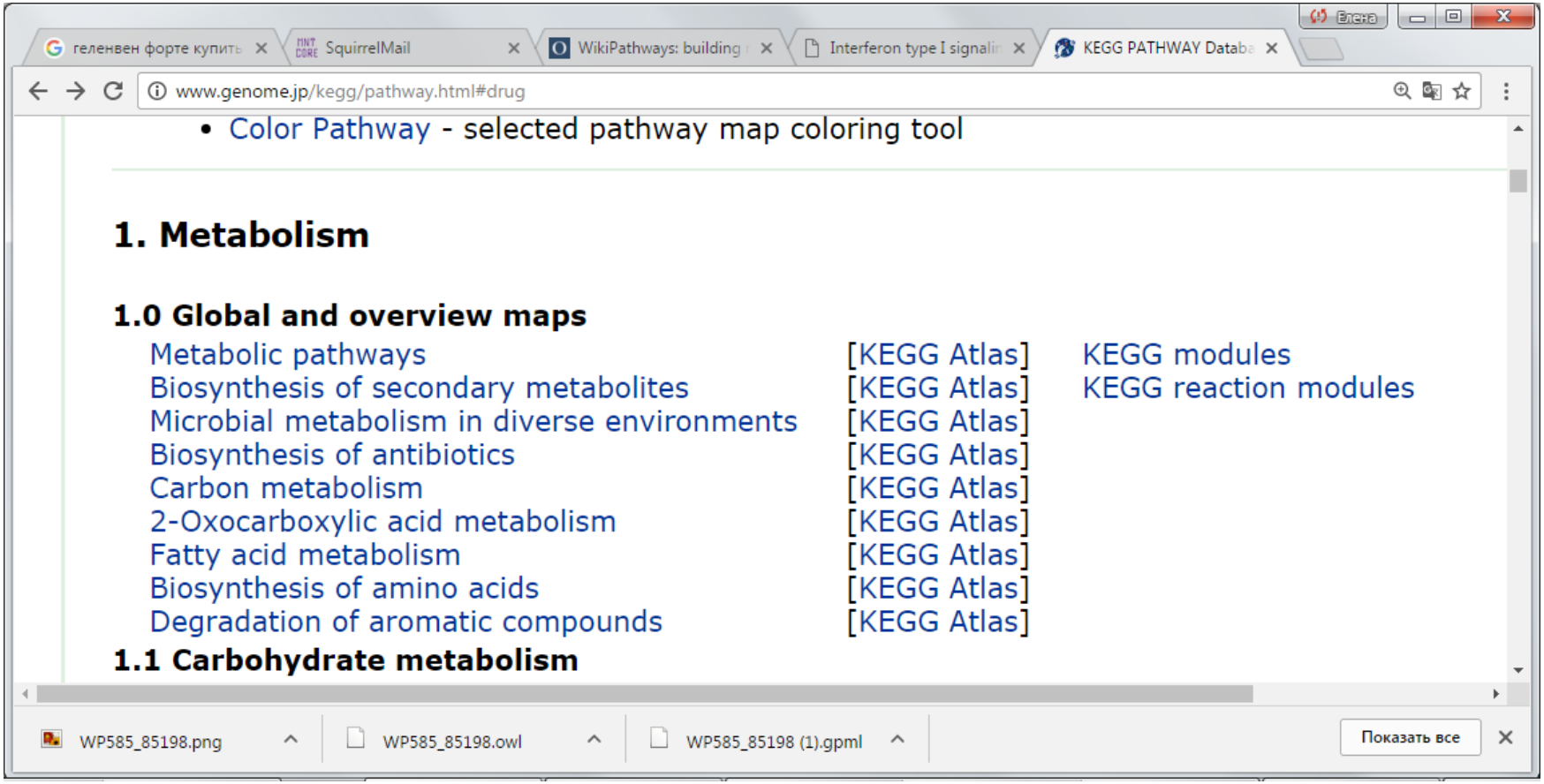
7. **Drug Development**

Pathway Mapping

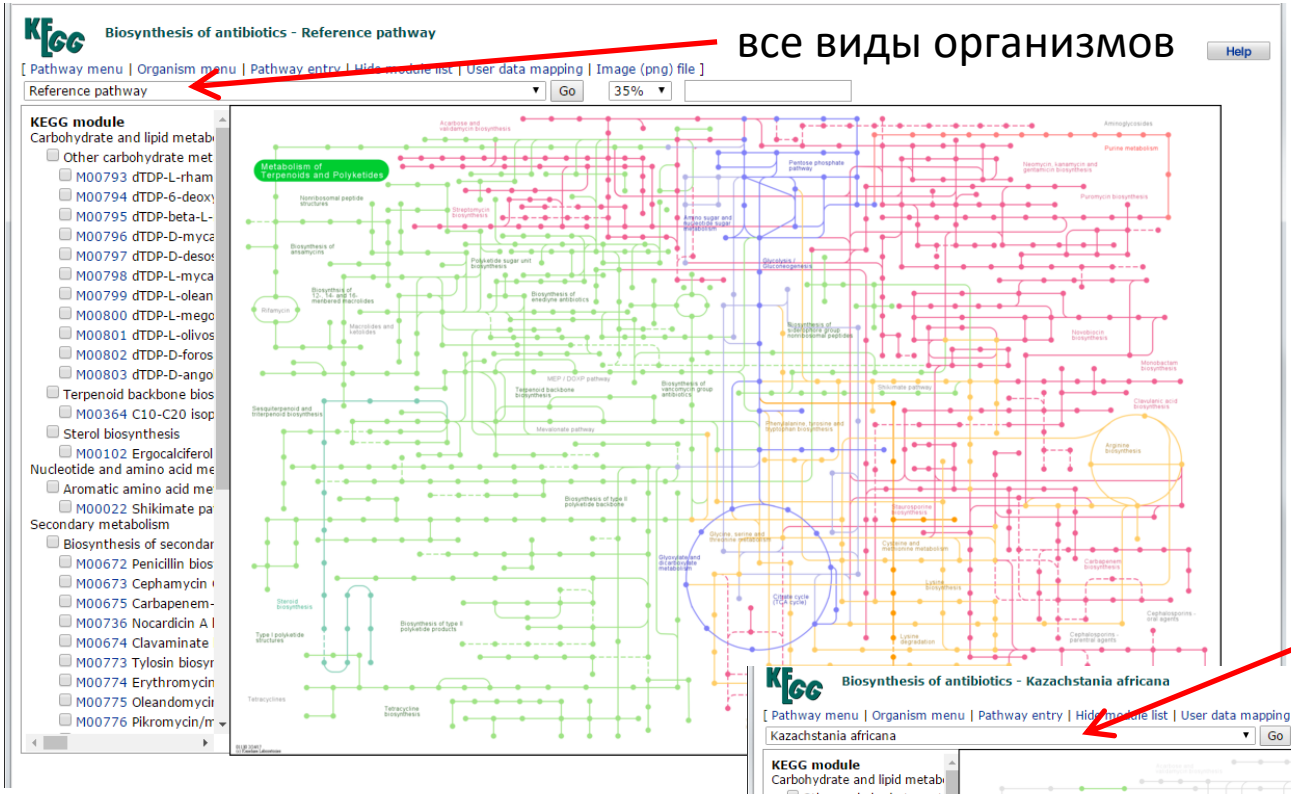
KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretaion of higher-level systemic functions.

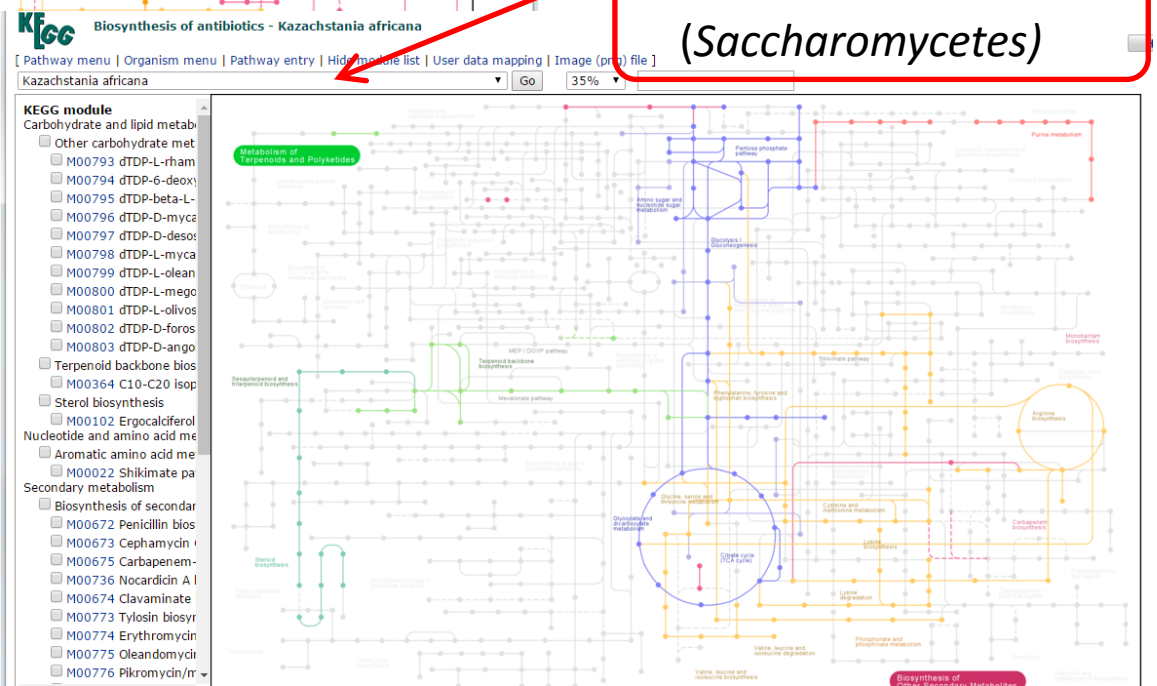
- Search Pathway - basic pathway mapping tool
- Search&Color Pathway - advanced pathway mapping tool
- Color Pathway - selected pathway map coloring tool

KEGG pathway: раздел , содержащий интегральные схемы (глобальные карты) метаболизма:

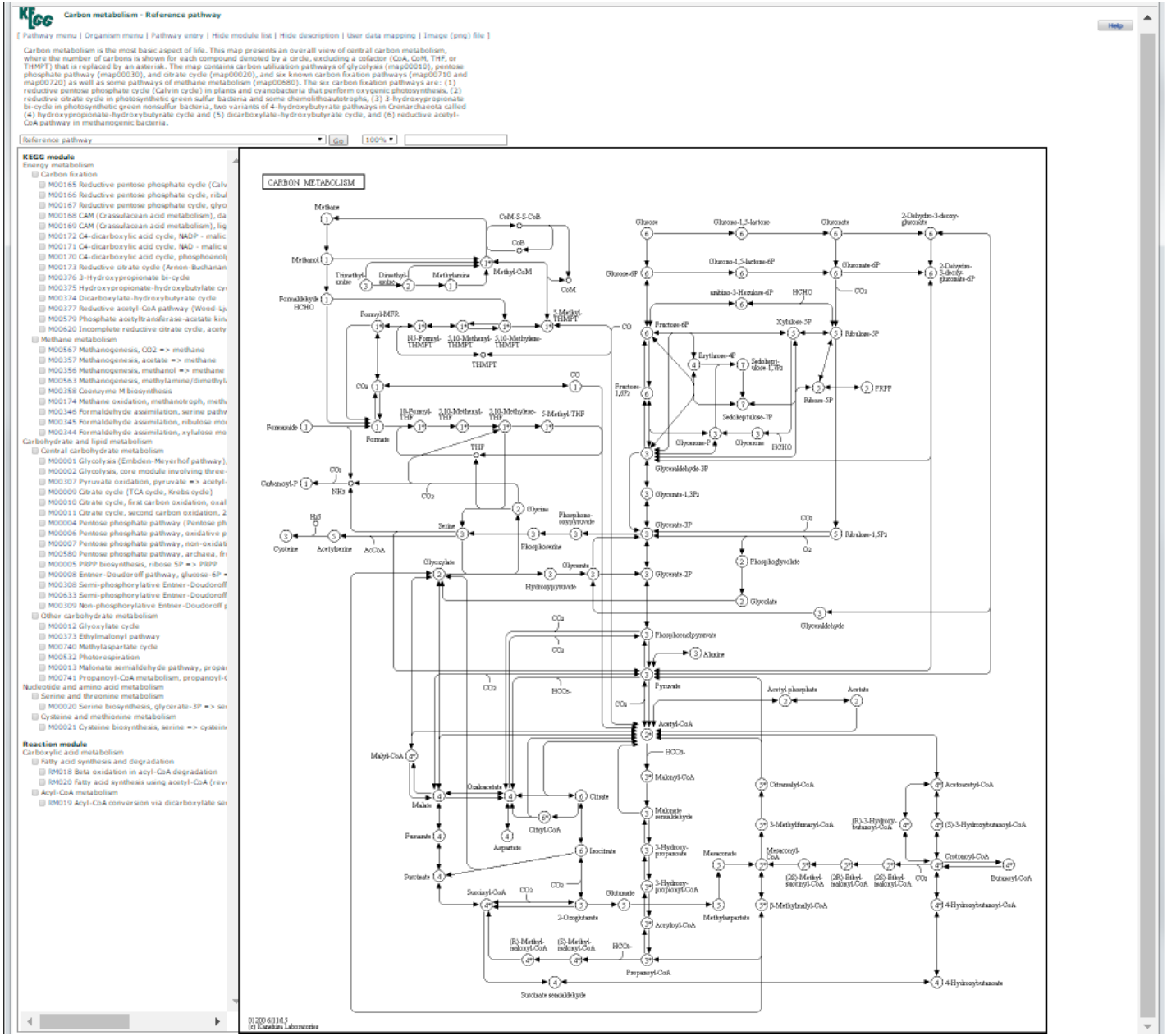


KEGG: Интегральная схема биосинтеза антибиотиков (все виды организмов)

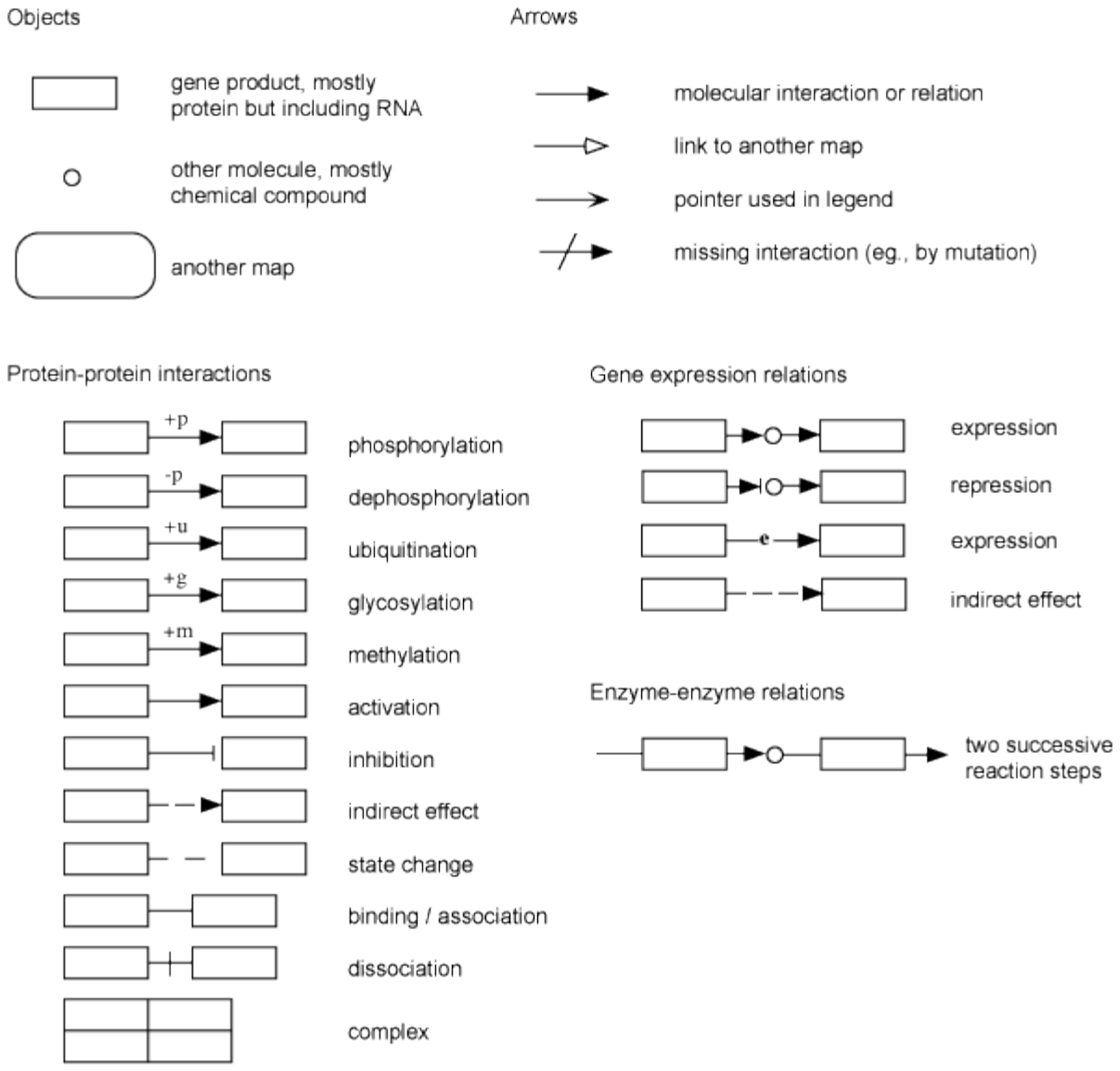




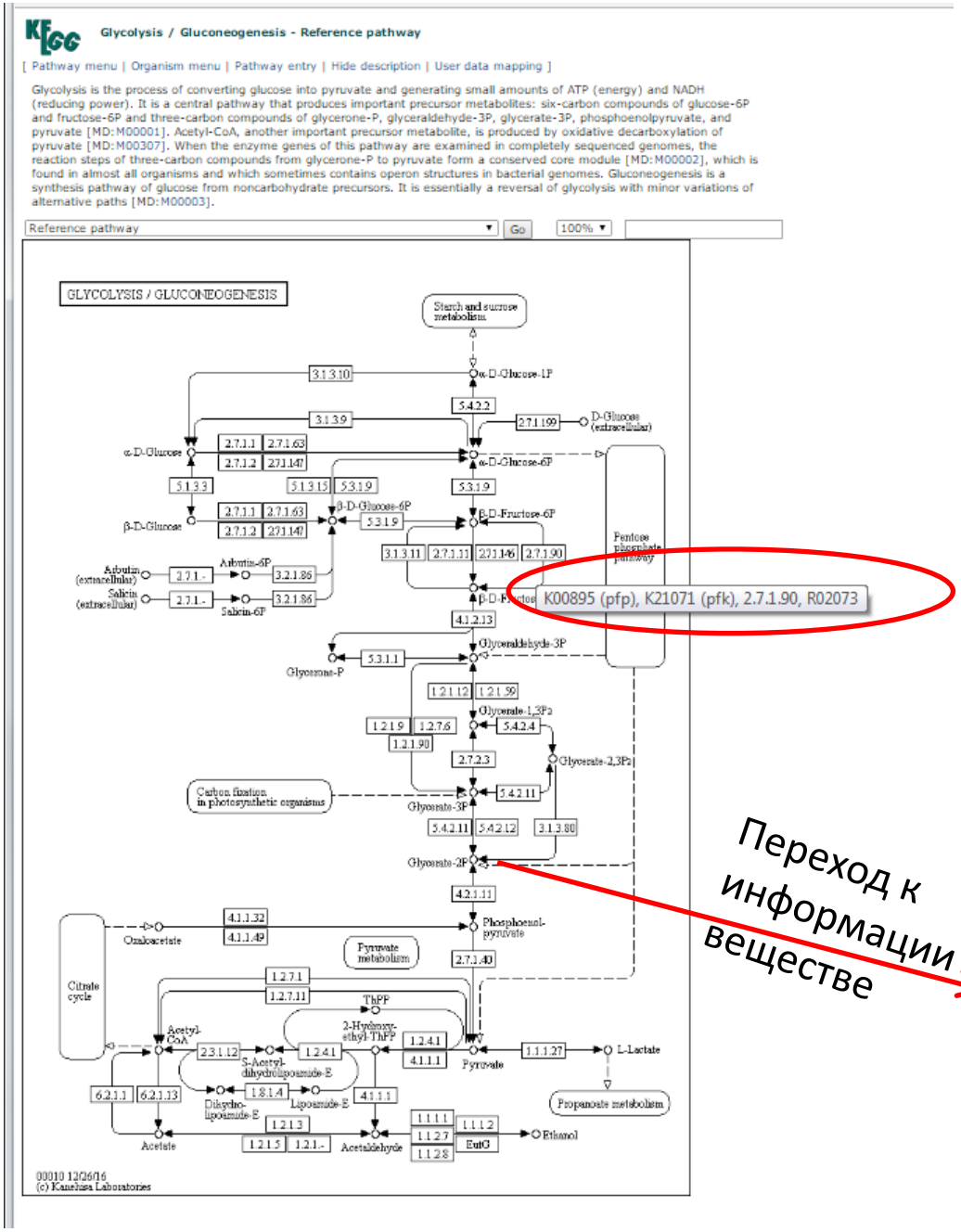
KEGG: Интегральная схема метаболизма углеводов (Carbon metabolism)



KEGG: условные обозначение

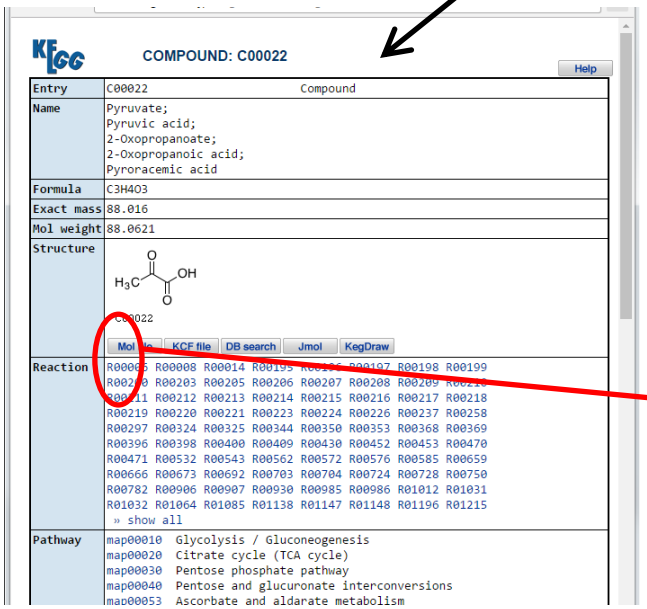


KEGG: диаграмма Glycolysis / Gluconeogenesis



Переход к информации о веществе

ВХОД “Pyruvate”
Разделе базы
“KEGG Chemical Universe”

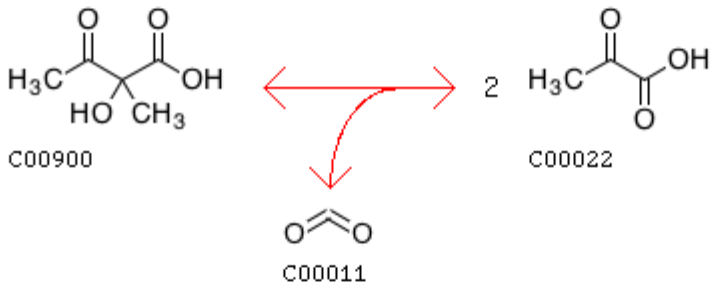


ВХОД “REACTION” в “**KEGG REACTION**”:
Одна из реакций с участием вещества «Pyruvate»

KEGG

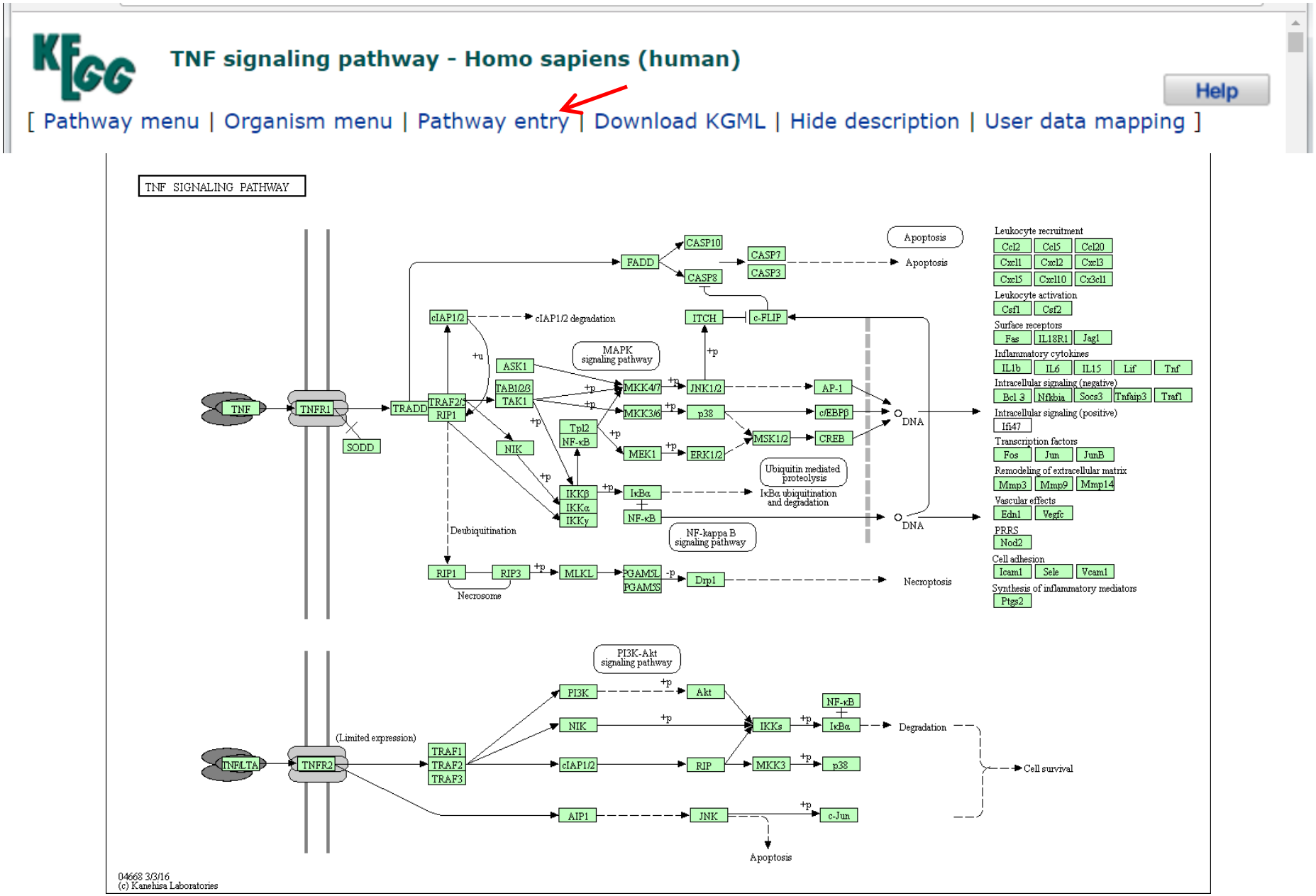
REACTION: R00006

Help

Entry	R00006 Reaction
Name	pyruvate:pyruvate acetaldehydetransferase (decarboxylating); 2-acetolactate pyruvate-lyase (carboxylating)
Definition	2-Acetolactate + CO2 <=> 2 Pyruvate
Equation	C00900 + C00011 <=> 2 C00022
	 C00900 C00011 2 C00022
Comment	TPP-dependent enzymatic reaction (R00014+R03050)
Reaction class	RC00106 C00022_C00900
Enzyme	2.2.1.6
Pathway	rn00770 Pantothenate and CoA biosynthesis
Orthology	K01652 acetolactate synthase I/II/III large subunit [EC:2.2.1.6] K01653 acetolactate synthase I/III small subunit [EC:2.2.1.6]

DBGET integrated database retrieval system

KEGG: схема сигнального пути , активируемого TNFa
(TNF signaling pathway - Homo sapiens (human))

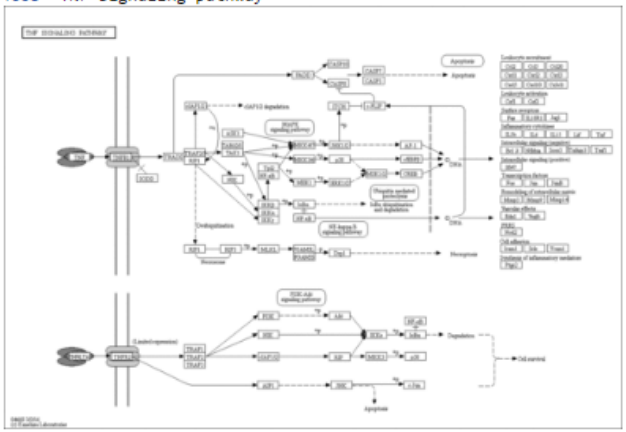


KEGG: информация из раздела Pathway entry для сигнального пути , активируемого TNFa **(TNF signaling pathway - Homo sapiens (human))**

KEGG

PATHWAY: hsa04668

Help

Entry	hsa04668 Pathway
Name	TNF signaling pathway - Homo sapiens (human)
Description	Tumor necrosis factor (TNF), as a critical cytokine, can induce a wide range of intracellular signal pathways including apoptosis and cell survival as well as inflammation and immunity. Activated TNF is assembled to a homotrimer and binds to its receptors (TNFR1, TNFR2) resulting in the trimerization of TNFR1 or TNFR2. TNFR1 is expressed by nearly all cells and is the major receptor for TNF (also called TNF-alpha). In contrast, TNFR2 is expressed in limited cells such as CD4 and CD8 T lymphocytes, endothelial cells, microglia, oligodendrocytes, neuron subtypes, cardiac myocytes, thymocytes and human mesenchymal stem cells. It is the receptor for both TNF and LTA (also called TNF-beta). Upon binding of the ligand, TNFR mediates the association of some adaptor proteins such as TRADD or TRAF2, which in turn initiate recruitment of signal transducers. TNFR1 signaling induces activation of many genes, primarily controlled by two distinct pathways, NF-kappa B pathway and the MAPK cascade, or apoptosis and necroptosis. TNFR2 signaling activates NF-kappa B pathway including PI3K-dependent NF-kappa B pathway and JNK pathway leading to survival.
Class	Environmental Information Processing; Signal transduction BRITE hierarchy
Pathway map	hsa04668 TNF signaling pathway  All organisms Ortholog table
Other DBs	BSID: 812256 GO: 0033209
Organism	Homo sapiens (human) [GN:hsa]
Gene	7124 TNF; tumor necrosis factor [KO:K03156] 7132 TNFRSF1A; TNF receptor superfamily member 1A [KO:K03158] 9530 BAG4; BCL2 associated athanogene 4 [KO:K09558] 8717 TRADD; TNFRSF1A associated via death domain [KO:K03171] 7186 TRAF2; TNF receptor associated factor 2 [KO:K03173] [EC:2.3.2.27] 7188 TRAF5; TNF receptor associated factor 5 [KO:K09849] 8737 RIPK1; receptor interacting serine/threonine kinase 1 [KO:K02861] [EC:2.7.11.1] 329 BIRC2; baculoviral IAP repeat containing 2 [KO:K16060] 330 BIRC3; baculoviral IAP repeat containing 3 [KO:K16060] 6885 MAP3K7; mitogen-activated protein kinase kinase kinase 7 [KO:K04427] [EC:2.7.11.25] 10454 TAB1; TGF-beta activated kinase 1 (MAP3K7) binding protein 1 [KO:K04403]

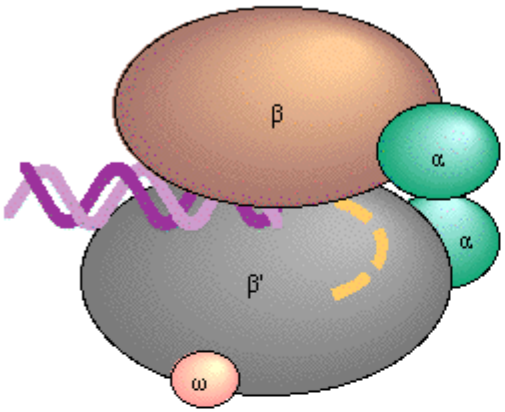
All links

Pathway (1)
BioSystems (1)
Genome (1)
KEGG GENOME (1)
Gene (108)
KEGG GENES (108)
All databases (110)

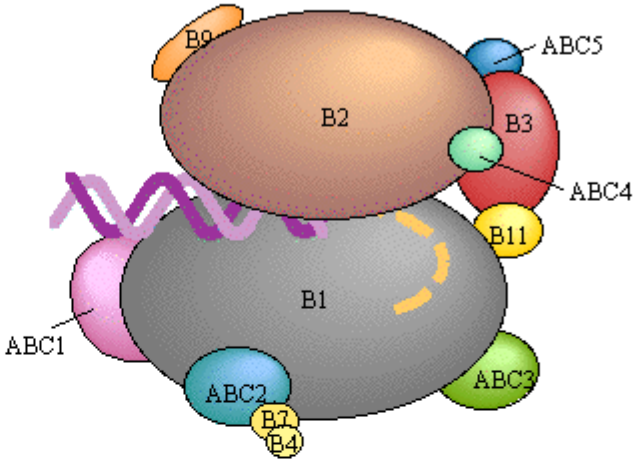
Download RDF

KEGG: диаграмма **RNA polymerase из раздела 2.1 Transcription**

RNA POLYMERASE



RNA polymerase (Thermus aquaticus)



RNA polymerase II (Saccharomyces cerevisiae)

Bacterial

β	α	ω	δ
β'			

Archaeal

B	D	F	H	K	E
A	G		N	L	P

Eukaryotic Pol II

Core subunits

B2	B3
B1	B11

Pol II specific subunits

B4	B7	B9

Pol I, II, and III common subunits

ABC1	ABC2	ABC3
ABC4	ABC5	

Eukaryotic Pol III

Core subunits

C2	AC2
C1	AC1

Pol III specific subunits

C3	C4	C11	
C25	C31	C34	C37

Eukaryotic Pol I

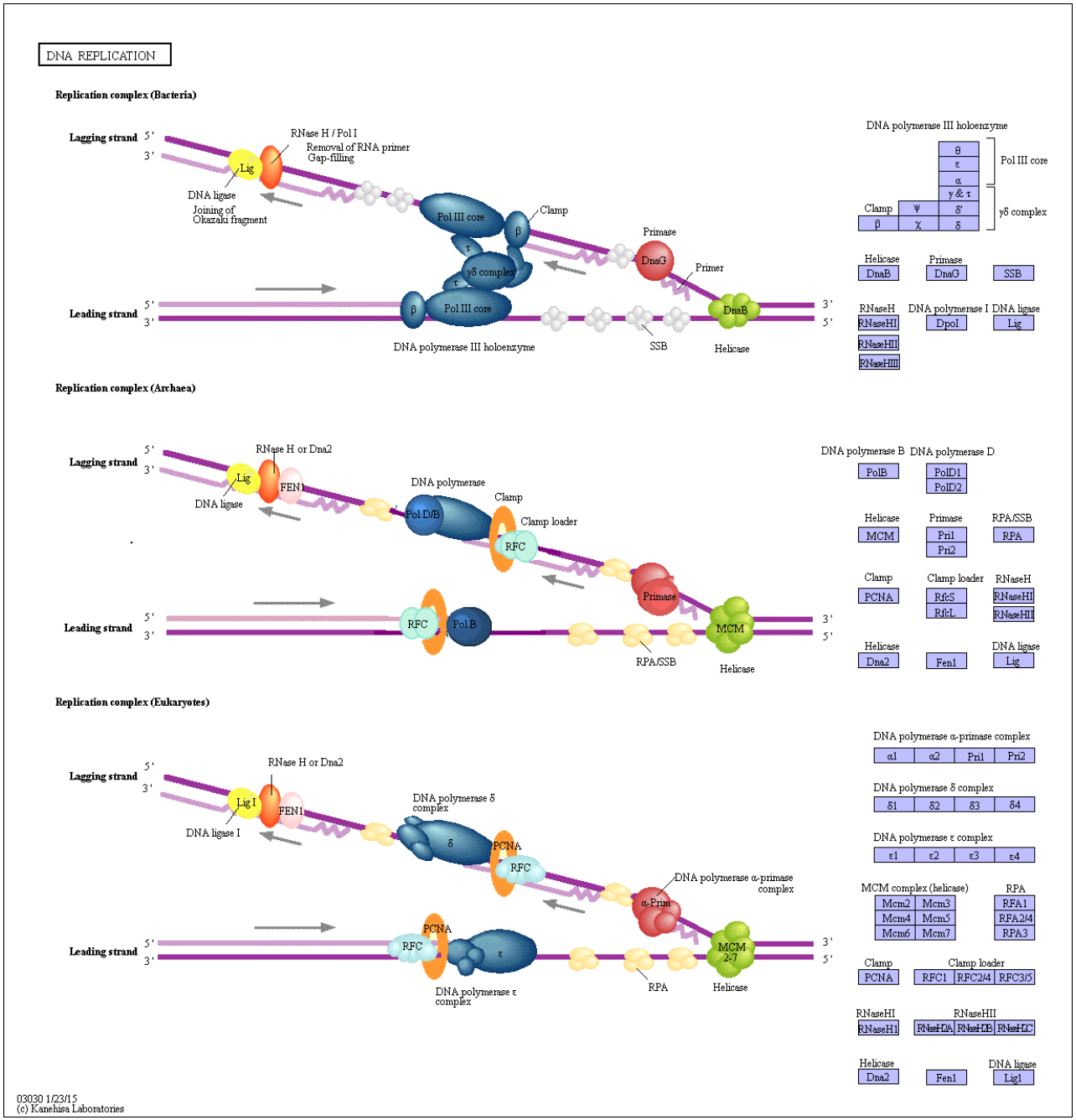
Core subunits

A2	AC2
A1	AC1

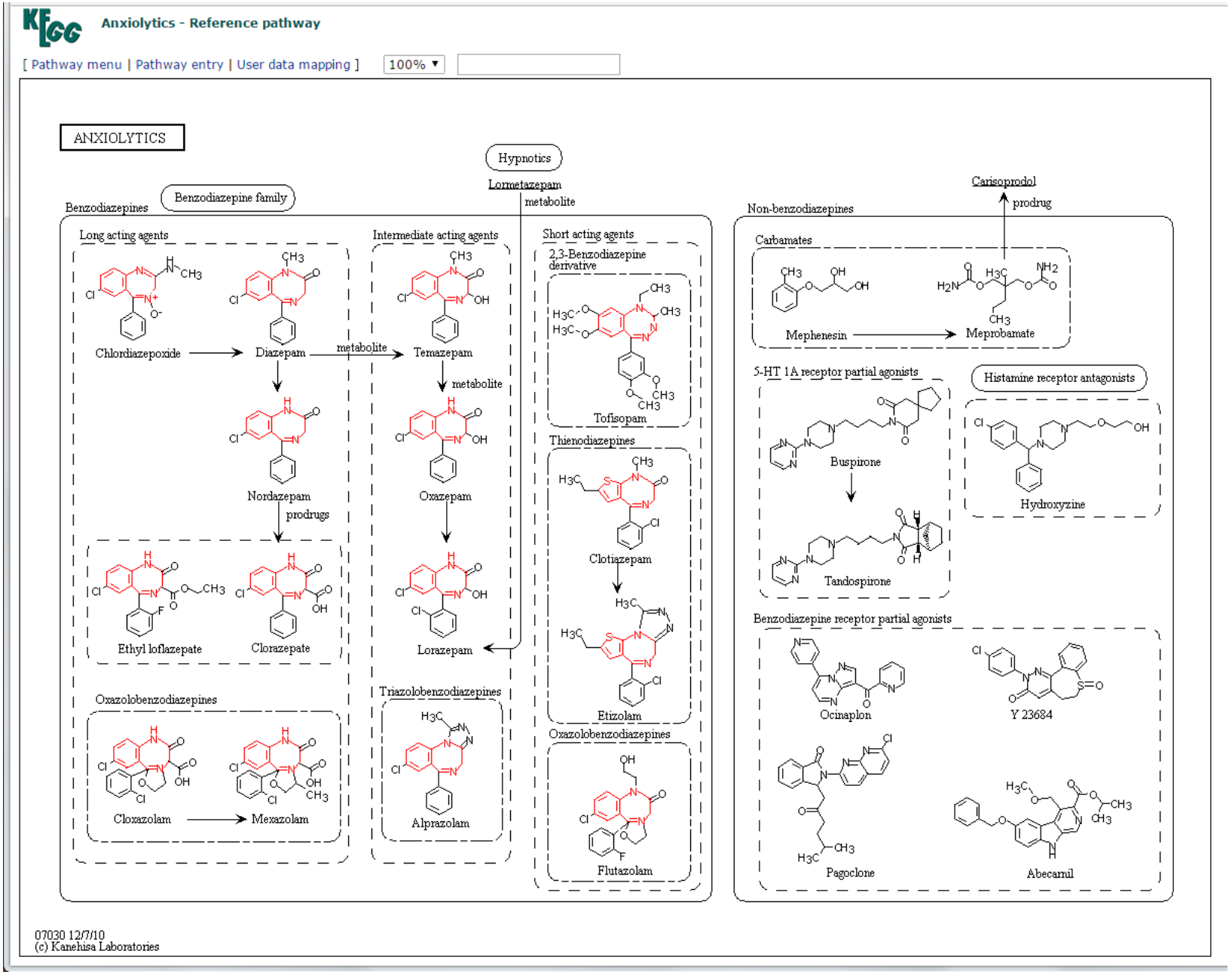
Pol I specific subunits

A12	A14	A34
A49	A43	

KEGG: диаграмма **DNA replication из раздела 2.4 Replication and repair**



KEGG: диаграмма **Anxiolytics из раздела 7.3 Chronology: Nervous system agents**



Анксиолитики (= транквилизаторы): вещества, снимающие тревогу, страх, усталость

Статистика базы KEGG pathway

leic Acids Res

Manually drawn KEGG reference pathway maps

Category	Type	Number of maps[a]
Metabolism	Global map	4
	Overview map	5
	Regular map	160
	Chemical structure transformation map	9
Genetic information processing	Regular map	22
Environmental information processing	Regular map	38
Cellular processes	Regular map	24
Organismal systems	Regular map	78
Human diseases	Regular map	81
Drug development	Drug structure map	75

[a]As of 1 October 2016.

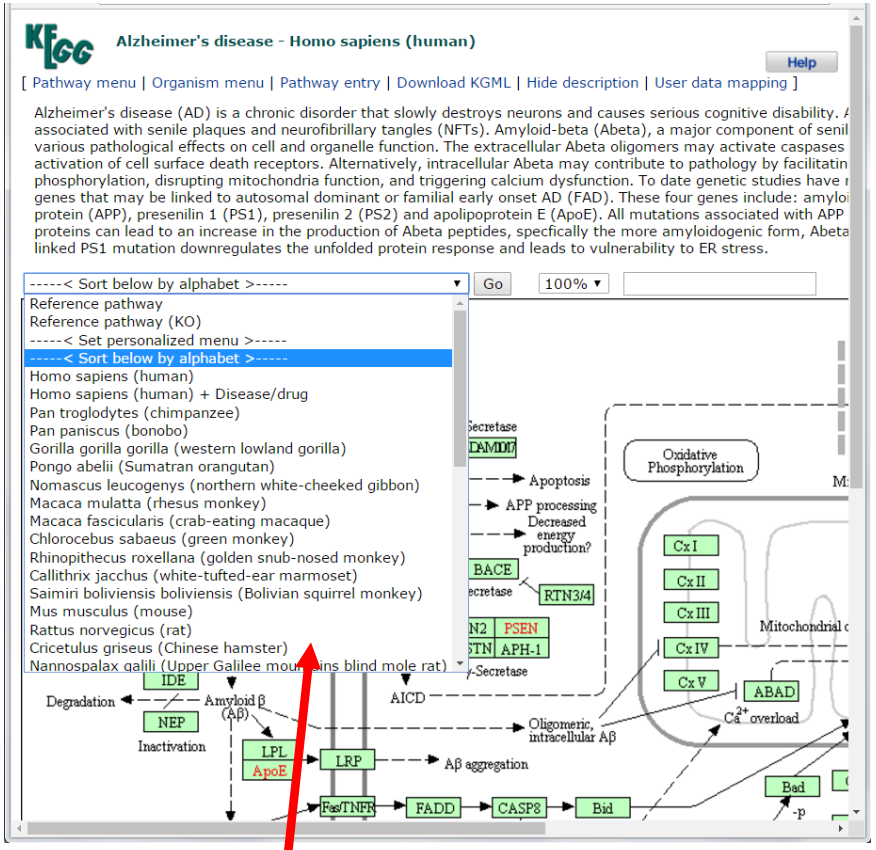
Σ = 496

Nucleic Acic

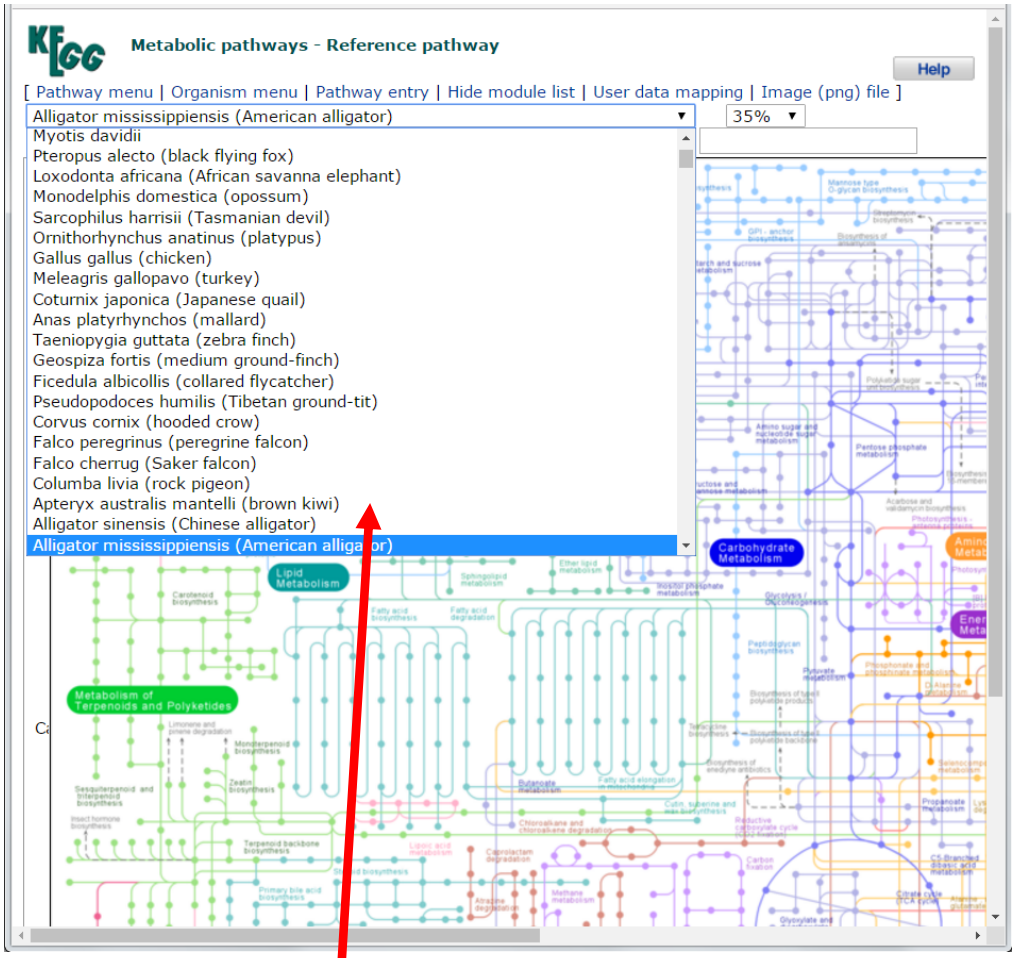
Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017 Jan 4;45(D1):D353-D361.

А сколько видов представлено в базе KEGG pathway?

- Ответ зависит от того, какой вход (pathway) мы рассматриваем.

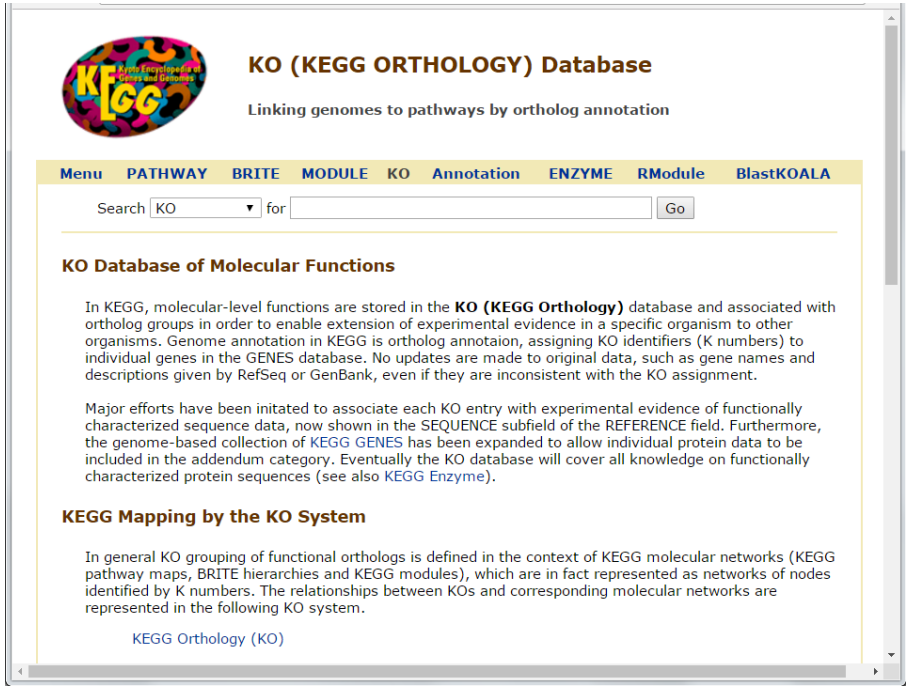


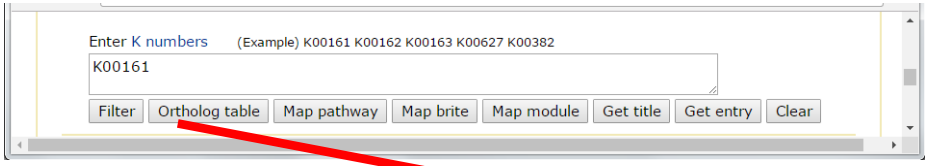
42 вида организмов



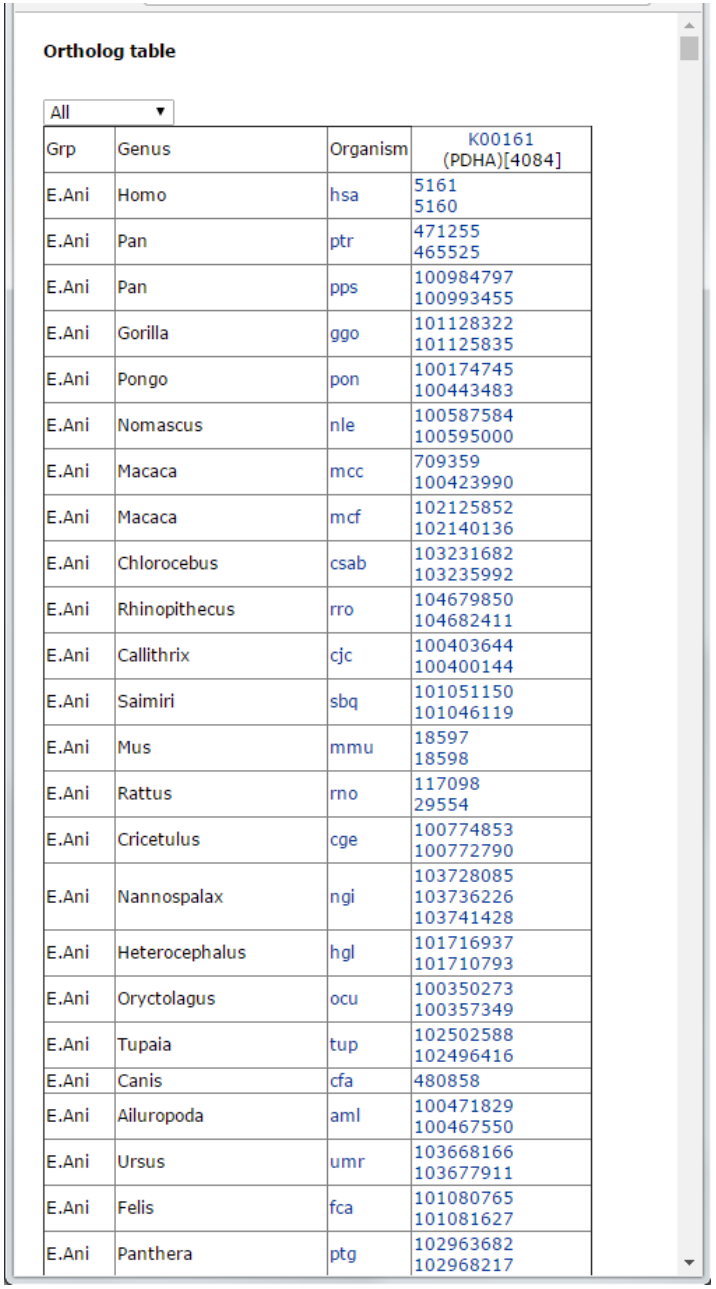
> 1000 видов организмов

KEGG ORTHOLOGY





Гены всех видов сгруппированы в группы ортологов. Каждая группа имеет индивидуальный идентификатор (вида K00161), через который можно вести поиск как по генам, так и по разделу KEGG Pathway

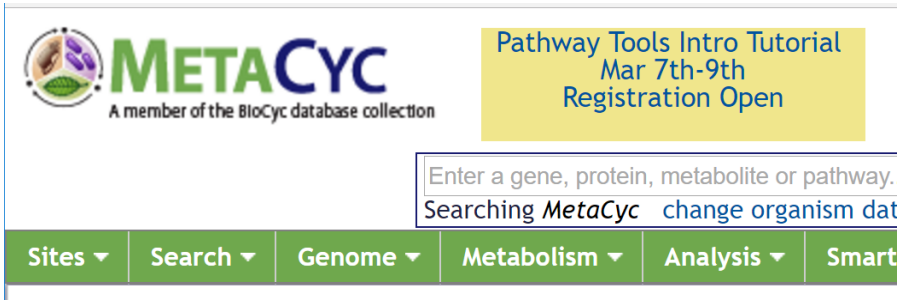


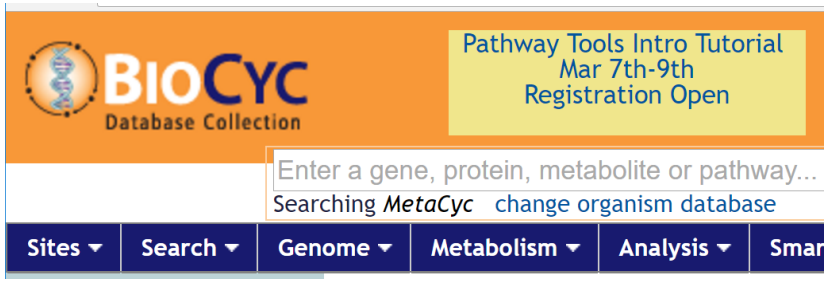
Ortholog table

All

Grp	Genus	Organism	K00161 (PDHA)[4084]
E.Ani	Homo	hsa	5161 5160
E.Ani	Pan	ptr	471255 465525
E.Ani	Pan	pps	100984797 100993455
E.Ani	Gorilla	ggo	101128322 101125835
E.Ani	Pongo	pon	100174745 100443483
E.Ani	Nomascus	nle	100587584 100595000
E.Ani	Macaca	mcc	709359 100423990
E.Ani	Macaca	mcf	102125852 102140136
E.Ani	Chlorocebus	csab	103231682 103235992
E.Ani	Rhinopithecus	rro	104679850 104682411
E.Ani	Callithrix	cjc	100403644 100400144
E.Ani	Saimiri	sbq	101051150 101046119
E.Ani	Mus	mmu	18597 18598
E.Ani	Rattus	rno	117098 29554
E.Ani	Cricetulus	cge	100774853 100772790
E.Ani	Nannospalax	ngi	103728085 103736226 103741428
E.Ani	Heterocephalus	hgl	101716937 101710793
E.Ani	Oryctolagus	ocu	100350273 100357349
E.Ani	Tupaia	tup	102502588 102496416
E.Ani	Canis	cfa	480858
E.Ani	Ailuropoda	aml	100471829 100467550
E.Ani	Ursus	umr	103668166 103677911
E.Ani	Felis	fca	101080765 101081627
E.Ani	Panthera	ptg	102963682 102968217

Семейство баз данных MetaCyc – BioCyc: метаболические пути





MetaCyc содержит более **2400** метаболических путей **из >46 000 публикаций** для **2816** видов организмов

BioCyc - Коллекция **9390** организм-специфичных Pathway/Genome Databases (PGDBs), каждая из которых содержит полный геном и **предсказанные** метаболические сети данного организма (включая описание метаболитов, ферменты, реакции, метаболичекие пути, предсказанные опероны, транспортные системы и фильтры, позволяющие получать информацию о метаболических путях)

Общий формат представления данных, поисковые системы, средства анализа

Базы развиваются с 2005 года (BioCyc) и 2001 года (MetaCyc)

Caspi R, et al., The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 2014;42:D459-71

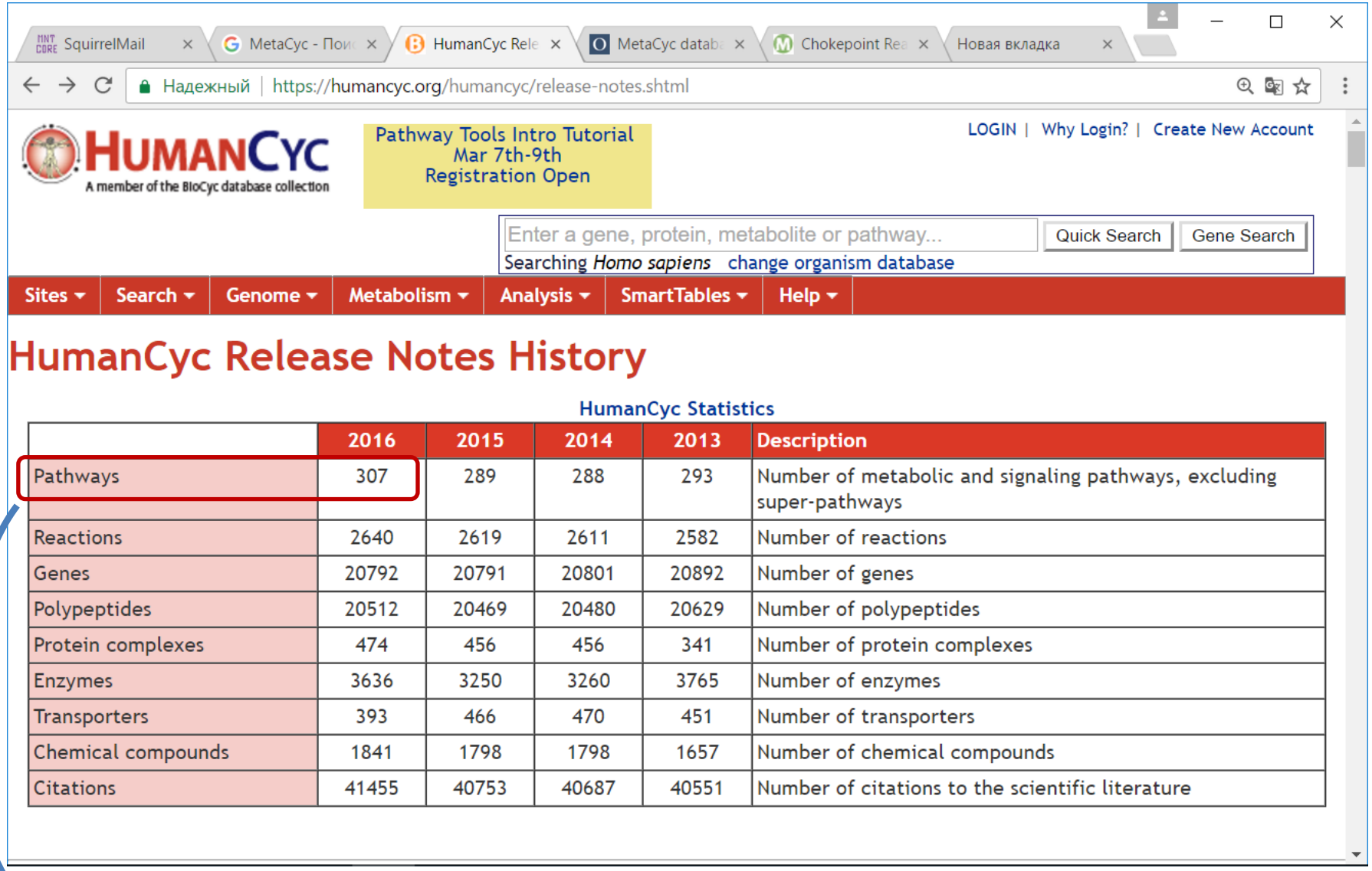
Caspi R et al., The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2016;44(D1):D471-80.

MetaCyc: список видов, для которых имеется 20 и более метаболических путей с экспериментально подтвержденными данными

Bacteria		Eukarya		Archaea	
Escherichia coli	329	*Arabidopsis thaliana*	335	*Methanocaldococcus jannaschii*	29
Pseudomonas aeruginosa	71	*Homo sapiens*	264	*Methanosarcina barkeri*	22
Bacillus subtilis	62	*Saccharomyces cerevisiae*	188	*Sulfolobus solfataricus*	21
Pseudomonas putida	51	*Rattus norvegicus*	83		
Salmonella typhimurium	41	*Glycine max*	62		
Pseudomonas fluorescens	32	*Solanum lycopersicum*	55		
Mycobacterium tuberculosis	31	*Pisum sativum*	55		
Klebsiella pneumoniae	29	*Mus musculus*	54		
Synechocystis sp. PCC 6803	27	*Zea mays*	48		
Enterobacter aerogenes	26	*Nicotiana tabacum*	46		
Agrobacterium tumefaciens	24	*Oryza sativa*	46		
		Solanum tuberosum	43		
		Catharanthus roseus	30		
		Spinacia oleraca	29		
		Hordeum vulgare	27		
		Triticum aestivum	25		
		Bos taurus	23		
		Petunia x hybrida	21		
		Sus scrofa	20		

Информационное содержание базы HumanCyc

https://humancyc.org/humancyc/release-notes.shtml



HumanCyc Release Notes History

HumanCyc Statistics

	2016	2015	2014	2013	Description
Pathways	307	289	288	293	Number of metabolic and signaling pathways, excluding super-pathways
Reactions	2640	2619	2611	2582	Number of reactions
Genes	20792	20791	20801	20892	Number of genes
Polypeptides	20512	20469	20480	20629	Number of polypeptides
Protein complexes	474	456	456	341	Number of protein complexes
Enzymes	3636	3250	3260	3765	Number of enzymes
Transporters	393	466	470	451	Number of transporters
Chemical compounds	1841	1798	1798	1657	Number of chemical compounds
Citations	41455	40753	40687	40551	Number of citations to the scientific literature

KEGG Pathway – 496 pathways

HumanCyc - Pathway: superoxide radicals degradation

HUMANCYC
A member of the BioCyc database collection

Please take 3 minute BioCyc survey

Welcome: Elena Ignatieva | Logout | Help | My preferences

Enter a gene, protein, metabolite or pathway... Quick Search Gene Search
Searching *Homo sapiens* change organism database

Sites | Search | Genome | Metabolism | Analysis | SmartTables | Help

Add to SmartTable

Homo sapiens Pathway: superoxide radicals degradation

Less Detail

Compound: H_2O
Synonyms: H2O, hydrogen oxide, water

If an enzyme name is shown in bold, there is experimental evidence for this

Synonyms: removal of superoxide radicals

Superclasses: Detoxification → Reactive Oxygen Species Degradation

Pathway Summary from MetaCyc:

General Background

All organisms living in an aerobic environment are exposed to reactive oxygen species (ROS) that are formed through metabolic processes and various environmental stresses such as drought, air pollutants, UV light and high light intensities, chilling temperatures and external chemicals [Van99, Alscher02]. For example, active oxygen species are produced during the β-oxidation of fatty acids or as a result of photorespiration in photosynthetic organisms [Frugoli96]. ROS such as superoxide and hydroxyl radicals as well as hydrogen peroxide can cause significant damage to proteins, nucleic acids and cell organelles.

Most of the aerobic organisms have developed defense systems to face oxidative stress and to scavenge oxidative radicals in the form of enzymes that can detoxify ROS, such as superoxide dismutase (SOD) and hydroperoxidase (CAT) [Beyer87]. For example, *Arabidopsis thaliana* , a member of the mustard family (oilseed plants), stores energy reserves preliminary as lipids that undergo β-oxidation during germination. The hydrogen peroxide that is produced during this metabolic process is detoxified by catalase. Another form of ROS, superoxide radicals, are by-products of aerobic electron transfer chains, and are disposed of by the action of superoxide dismutase. Since ROS can be found in any compartment of the eukaryotic cell, organisms have developed small gene families encoding for several SOD and CAT enzymes that operate in the various cell compartments [Kliebenstein98, McClung97].

About This Pathway

SODs represent the first line of defense against ROS, converting superoxide radicals to hydrogen peroxide and water. SODs are iron-dependent, manganese-dependent and copper/zinc-dependent SODs,

https://humancyc.org/compound?orgid=HUMAN&id=WATER

HumanCyc - Pathway: superoxide radicals degradation (продолжение)

show operations

About This Pathway

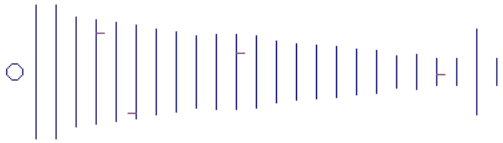
SODs represent the first line of defense against ROS, converting superoxide radicals to hydrogen peroxide and water. SODs are differentiated with regard to their metal cofactor. There are iron-dependent, manganese-dependent and copper/zinc-dependent SODs, which differ not only in their metal cofactor but also in their subcellular location. In plants, FeSODs are located in chloroplasts and are regarded the most ancient SOD group. MnSODs are found in the mitochondrion and the peroxisome and are structurally very similar to FeSODs. The last group, the Cu-ZnSODs operates in chloroplasts, the cytosol and even the extracellular space. They are structurally very different from the other two SOD groups because of the different electrical properties of copper in comparison to iron or manganese, which resulted in a major structural change in the protein [Alscher02].

To date seven SODs have been identified in *Arabidopsis thaliana* , three of them iron-dependent, three having copper as metal cofactor and one manganese-dependent SOD [Hindges92, Van90, Kliebenstein98]. It has been demonstrated that a copper-chaperone (AtCCS, At1g12520) is crucial for the activation of all three Cu/Zn-dependent SODs in this organism. The SOD holoenzyme usually constitutes either a homodimer or a homotetramer. However, the exact composition of the SODs in Arabidopsis is currently not known and remains to be verified (here displayed as polypeptides).

Catalase is second in the defense line against active oxygen, converting hydrogen peroxide into water and oxygen. Three genes encoding subunits of catalase and at least 6 catalase isoenzymes have been identified in Arabidopsis so far [Zhong94, McClung97, Frugoli96, Zhong96]. Besides their implication in detoxifying ROS, catalases are thought to play a role in the signal transduction pathway in plants leading to the development of SAR (systemic aquired resistance) [Jones94]. The functional protein of catalase is a tetramer but the question whether it exists as homo- or heterotetramer of different subunits remains to be investigated.

Superpathways: reactive oxygen species degradation

Locations of Mapped Genes:



Pathway Evidence Glyph:

This organism is in the expected taxonomic range for this pathway.

Key to pathway glyph edge colors:

An enzyme catalyzing this reaction is present in this organism
The reaction is unique to this pathway in MetaCyc

Credits:

Created in MetaCyc 07-Dec-1994 by Riley M , Marine Biological Laboratory
Reviewed in MetaCyc 30-Nov-2006 by Foerster H , The Arabidopsis Information Resource
Revised in MetaCyc 20-Feb-2009 by Caspi R , SRI International
Imported from MetaCyc 01-Nov-2016 by Caspi R , SRI International

References

Alscher02: Alscher RG, Erturk N, Heath LS (2002). **"Role of superoxide dismutases (SODs) in controlling oxidative stress in plants."** J Exp Bot 53(372);1331-41. PMID: 11997379

Beyer87: Beyer WF Jr, Fridovich I (1987). **"Catalases-with and without heme."** In MG Simic, KA Taylor, JF Ward, C Von Sonntag, eds, Oxygen Radicals in Biology and Medicine. Plenum, New York, 651-661.

HumanCyc –Pathway: superoxide radicals degradation – информация о ферменте SOD

show operations

Add to SmartTable

gene	enzyme
SOD1	**superoxide dismutase [Cu-Zn]**
Homo sapiens	

Synonyms	IPOA, ALS1, ALS, superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult)), Cu/Zn superoxide dismutase, Superoxide dismutase-1, soluble		
Accession IDs	HS06899 (HumanCyc) P00441 (UniProt)	Length	8905 bp
		Map Position	[29,692,513 -> 29,701,417] (61.86 centisomes) on Chromosome 21 View in Genome Browser
		Location	cytosol
Reaction	2 superoxide + 2 H^+ → hydrogen peroxide + oxygen		
Pathways	superoxide radicals degradation reactive oxygen species degradation		
Evidence	Assay of protein purified to homogeneity from a heterlogous host [Kajihara88]		

Summary | GO Terms (8) | Essentiality | Reactions (1) | Protein Features | Gene Context | References | Show All

Summary

Superoxide dismutases are metalloproteins which destroy radicals normally produced within cells that are toxic to biological systems. They catalyze the dismutation of superoxide to oxygen and hydrogen peroxide [Kajihara88].

Superoxide dismutases are classified according to the metal ion, Zn, Mn or Cu, which are a required cofactor for enzymic activity. There are three known superoxide dismutases in humans, superoxide dismutase [Cu-Zn], superoxide dismutase [Mn] and extracellular superoxide dismutase [Cu-Zn]. Human superoxide dismutase [Cu-Zn] is a non-disulfide linked homodimer which binds one Cu and one Zn per subunit and is found primarily in the cytosol [Banci98][Arnesano04].

Human superoxide dismutase [Cu-Zn] is succinylated and succinylation decreases enzyme activity. NAD-dependent deacetylase sirtuin-5 binds to superoxide dismutase [Cu-Zn], desuccinylating and activating the enzyme. Mutations of the superoxide dismutase [Cu-Zn] succinylation site inhibits the growth of lung tumor cells, revealing a role for succinylation and desuccinylation of this enzyme in the growth of certain tumors [Lin13].

Mutations in the *SOD1* gene give rise to familial amyotrophic lateral sclerosis (ALS) also known as Lou Gehrig's disease [Hart98][Nakano94]. Familial amyotrophic lateral sclerosis is a neurodegenerative disorder affecting upper motor neurons in the brain and lower motor neurons in the brain stem and spinal cord, resulting in fatal paralysis.

Experiments *in vitro* have shown that the loss of zinc from superoxide dismutase [Cu-Zn] induces apoptosis in cultured motor neurons mediated by nitric oxide. Both familial and sporadic ALS may involve an oxidative mechanism requiring nitric oxide [Estevez99].

Additional Citations: [Levanon85, Sherman83]

Subunit Composition	$[SOD1]_2$

Unification Links

Ensembl	ENSG00000142168
Entrez	AAB05661, AAB05662, AAB59492, AAB59626, AAB59627, AAC41773, AAH01034, CAA25915, CAA25916, CAA25917, CAA25918, CAA25919, CAA26182, CAA64520
Entrez-gene	6647
Entrez-Nucleotide	BC001034, K00065, L44139, X02317
GeneCards	SOD1
MOPED	P00441
OMIM	105400, 147450, 158700
PDB	1AZV, 1BA9, 1MFM, 1SOS, 1SPD
RefSeq	NM_000454, NP_000445
UCSC Human Genome	NM_000454
UniGene	75428
UniProt	P00441

Gene-Reaction Schematic

Expand/Contract the Schematic connections: Expand?

Credits:
Created 09-Jun-2014 by SRI International

EcoCyc : информационное содержание

Data type	Number (Release 20.1)
Genes	4505
Gene products covered by a mini-review	3884
Gene products with GO terms with EXP evidence	3350
Enzymes	1567
Metabolic reactions	1913
Compounds	2699
Transporters	282
Transport reactions	485
Transported substrates	338
Transcription factors	204
Regulatory interactions	6399

EcoCyc as Keseler et al. (2017), "EcoCyc: reflecting new knowledge about Escherichia coli K-12", Nucleic Acids Research 45:D543-50.

EcoCyc : статистика по данным, относящимся к регуляции транскрипции

Data type	Total	New
Transcription Unit	3553	95
Promoter	3841	73
Terminator	283	31
Transcription Factor	205	14
Transcription Factor Binding Site	2836	199
Regulatory Interaction	3374	183

EcoCyc as Keseler et al. (2017), "EcoCyc: reflecting new knowledge about Escherichia coli K-12", Nucleic Acids Research 45:D543-50.

Представление данных в EcoCyc: результат поиска по названию гена aceB

EcoCyc
A member of the BioCyc database collection

Please take 3 minute BioCyc survey

LOGIN | Why Login? | Create New Account

Enter a gene, protein, metabolite or pathway... Quick Search Gene Search
Searching *Escherichia coli K-12 substr. MG1655 (EcoCyc)* change organism database

Sites | Search | Genome | Metabolism | Analysis | SmartTables | Help

show operations

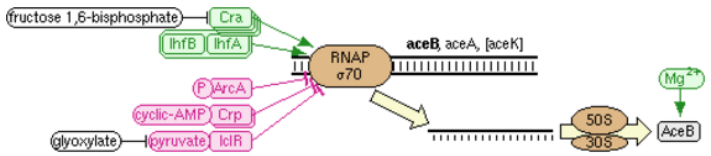
log in to add to SmartTable.

gene **aceB** enzyme **malate synthase A**
Escherichia coli K-12 substr. MG1655

Synonym	mas		
Accession IDs	EG10023 (EcoCyc) b4014 ECK4006 P08997 (UniProt)	Length	1602 bp / 533 aa
		Map Position	[4,215,478 -> 4,217,079] (90.82 centisomes, 327°) View in Genome Browser
		Location	cytosol
Reaction	acetyl-CoA + glyoxylate + H_2O → (S)-malate + coenzyme A + H^+		
Pathway	glyoxylate cycle		
Evidence	Assay of protein purified to homogeneity from its native host [Lohman08] Assay of unpurified protein expressed in its native host [Chung88]		

Summary | GO Terms (7) | Essentiality | Reactions (1) | Protein Features | Operons | References | Show All

Regulation Summary Diagram



Summary

There are two isozymes of malate synthase in *E. coli* [Falmagne65, Molina94]. Malate synthase A encoded by the aceB gene is a key enzyme in the glyoxylate cycle. It metabolizes glyoxylate formed in the dissimilation of acetate. The relatively more studied isozyme malate synthase G encoded by the glcB gene is responsible for almost all of the malate synthase activity in cells metabolizing glyoxylate formed during growth on glycolate [Molina94].

Malate synthase A catalyzes the irreversible condensation of acetyl-CoA with glyoxylate to produce (S)-malate and coenzyme A. The formation of (S)-malate is a key reaction in the glyoxylate cycle. This cycle is similar to the TCA cycle (see TCA cycle I (prokaryotic)) but it bypasses the TCA cycle reactions that lead to a loss of CO_2, thus providing TCA cycle intermediates for cell carbon biosynthesis (see superpathway of glyoxylate bypass and TCA). The glyoxylate cycle has been extensively studied in connection with growth on acetate which is used in the synthesis of these biosynthetic precursors.

E. coli malate synthase A has been structurally characterized [Lohman08]. Its encoding gene *aceB* is located in the *aceBAK* operon which is

Unification Links

ASAP	ABE-0013125
CGSC	1051
EchoBASE	EB0022
EcoGene	EG10023
EcoliWiki	b4014
ModBase	P08997
OU-Microarray	b4014
PortEco	aceB
PR	PRO_000022037

Представление данных в EcoCyc: результат поиска по названию гена aceB (продолжение)

lacks the α/β domain found in isoform G. However when substrate is bound, the two active sites are nearly identical and inhibitors bind with similar affinities [Lohman08].

A series of vectors inducibly expressing paired-terminus antisense RNAs was constructed to silence central carbon metabolism in host *E. coli* K-12 MG1655. A vector that silenced *aceB* mRNA at 98% efficacy resulted in a defect in carbon catabolite repression [Nakashima14].

Review: [Cortay89]

Additional Citations: [Byrne88, Byrne88a, LaPorte85]

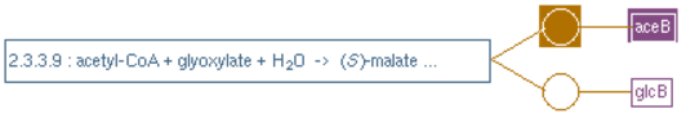
Molecular Weight of Polypeptide	60.274 kD (from nucleotide sequence)

show operations

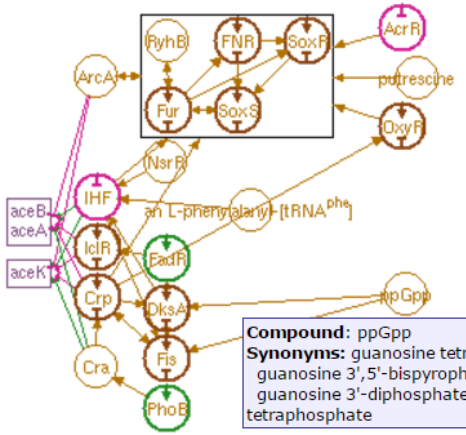
Relationship Links

InterPro: In-Family	IPR001465, IPR006252, IPR011076, IPR019830
Panther: In-Family	PTHR21631:SF1
PDB: Structure	3CUZ, 3CV1, 3CV2
Pfam: In-Family	PF01274
Prosite: In-Family	PS00510

Gene-Reaction Schematic



Genetic Regulation Schematic



Compound: ppGpp
Synonyms: guanosine tetraphosphate, guanosine 5'-diphosphate,3'-diphosphate, guanosine 3',5'-bispyrophosphate, guanosine 3',5'-bis(diphosphate), guanosine 3'-diphosphate 5'-diphosphate, magic spot, guanosine-5',3'-tetraphosphate

Регуляторная генная сеть !!!!

History:
10/20/97 Gene b4014 from Blattner lab Genbank (v. M52) entry merged into EcoCyc gene EG10023; confirmed by SwissProt match.

Credits:
Last-Curated 08-Jan-2016 by Fulcher C , SRI International

Report Errors or Provide Feedback
Page generated by Pathway Tools version 20.5 (software by SRI International) on Sun Mar 5, 2017, BIOCYC17A.
EcoCyc version 20.5.

https://ecocyc.org/compound?orgid=ECOLI&id=GUANOSINE-5DP-3DP

BioCyc: скачать данные можно, только купив лицензию

DOWNLOAD BIOCYC DATABASES AND PATHWAY TOOLS SOFTWARE

We provide two types of downloadable materials for the BioCyc databases and Pathway Tools software.

Note that the BioCyc web-based SmartTables facility can **save you significant time** in answering large-scale data analysis questions, and is significantly easier to use than is downloading and parsing BioCyc files. See the SmartTables menu.

Download BioCyc Data Files

We provide the BioCyc databases (such as EcoCyc and MetaCyc) as collections of data files in several alternative formats including the following.

Due to BioCyc moving to a subscription model in 2016, access to BioCyc DBs other than EcoCyc or MetaCyc requires purchase of a subscription (with the exception that older versions of BioCyc are freely available).

- BioPAX format
- Pathway Tools attribute-value format
- Pathway Tools tabular format
- SBML format
- Gene Ontology annotations (EcoCyc only): [Download from GO Web site]

Click here for an exact listing of files provided and their formats.

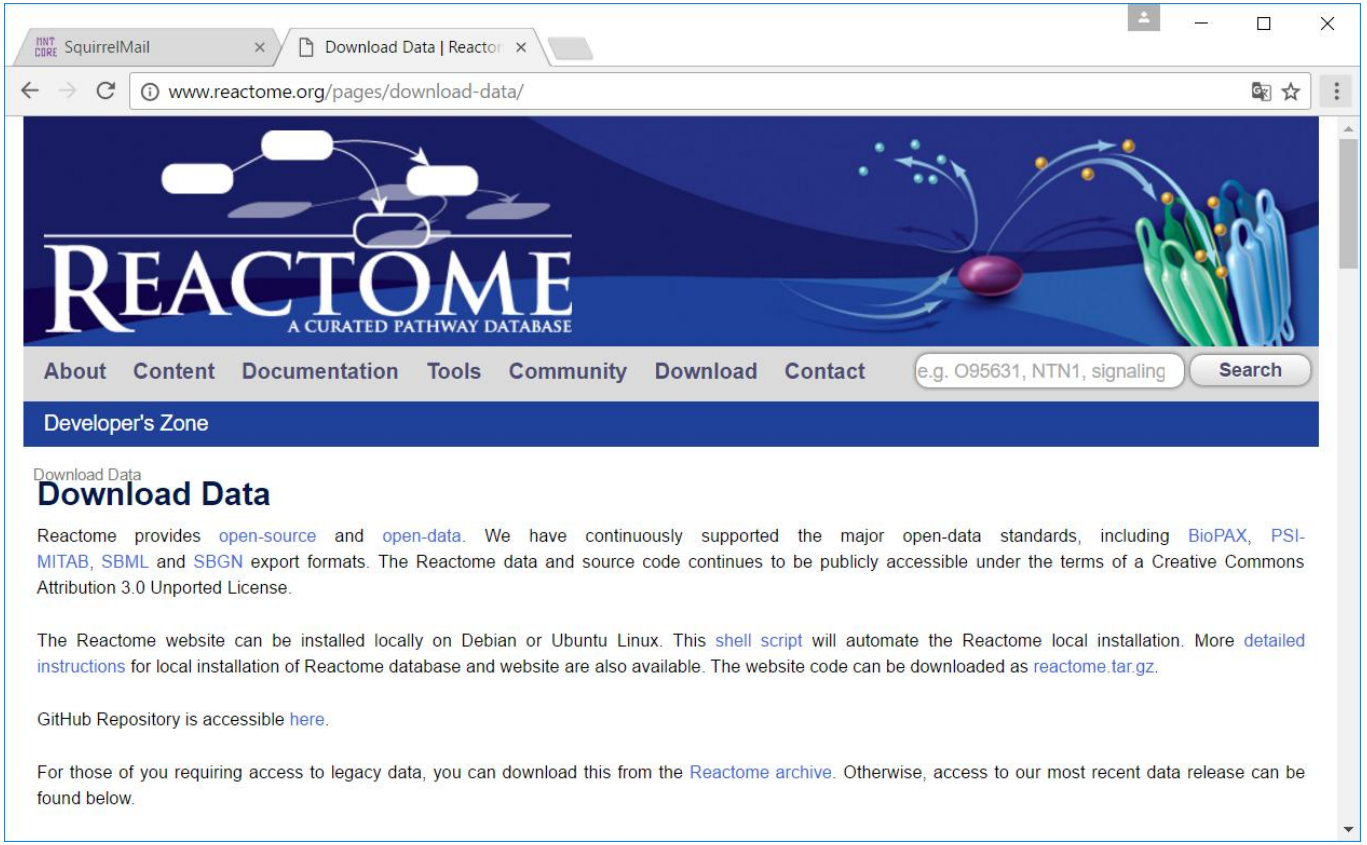
Programmatic Access to BioCyc Data

We provide several APIs for accessing BioCyc data.

БАЗА ДАННЫХ REACTOME – метаболические, сигнальные пути, схемы процессов

http://www.reactome.org/

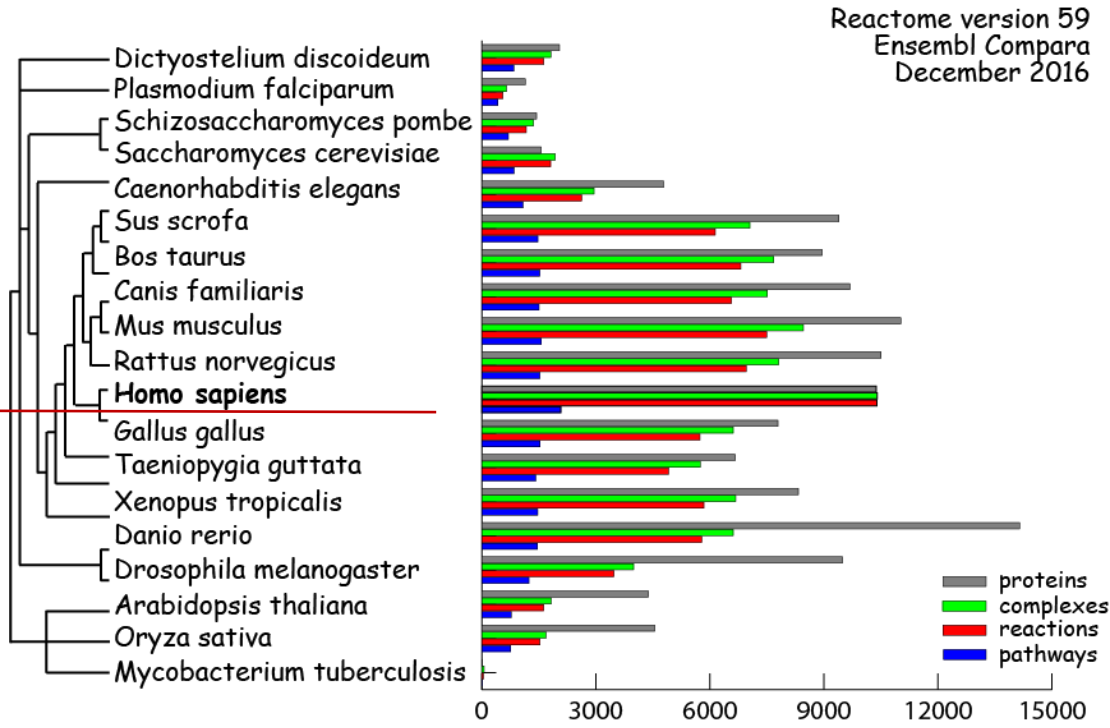
Данные находятся в открытом доступе !!



Reactome provides open-source and open-data. We have continuously supported the major open-data standards, including BioPAX, PSI-MITAB, SBML and SBGN export formats. The Reactome data and source code continues to be publicly accessible under the terms of a Creative Commons Attribution 3.0 Unported License.

БАЗА ДАННЫХ REACTOME(Version 59): статистика

Species	PROTEINS	COMPLE-XES	REAC-TIONS	PATHWAYS
D. discoideum	2042	1815	1631	841
P. falciparum	1144	648	555	424
S. pombe	1440	1362	1167	695
S. cerevisiae	1558	1932	1810	850
C. elegans	4780	2954	2627	1084
S. scrofa	9394	7050	6141	1473
B. taurus	8951	7684	6813	1526
C. familiaris	9683	7514	6566	1504
M. musculus	11020	8459	7491	1568
R. norvegicus	10506	7810	6957	1528
*H. sapiens	10374	10399	10391	2080
G. gallus	7790	6615	5737	1533
T. guttata	6666	5746	4915	1419
X. tropicalis	8328	6677	5841	1468
D. rerio	14148	6617	5784	1467
D. melanogaster	9487	3996	3472	1239
A. thaliana	4381	1821	1624	772
O. sativa	4550	1690	1530	761
M. tuberculosis	13	58	40	12

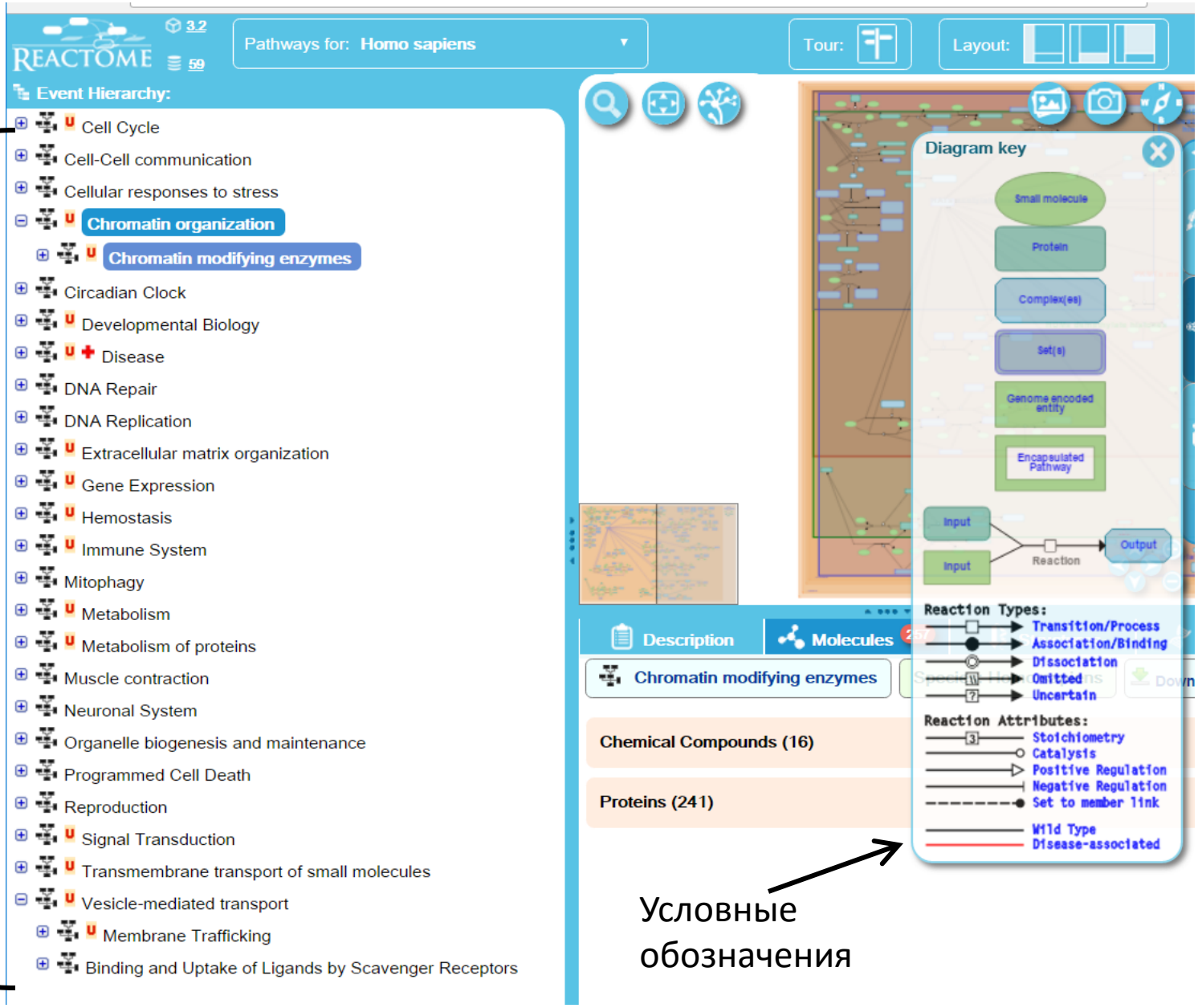


---

19 видов организмов

БАЗА ДАННЫХ REACTOME: интерфейс

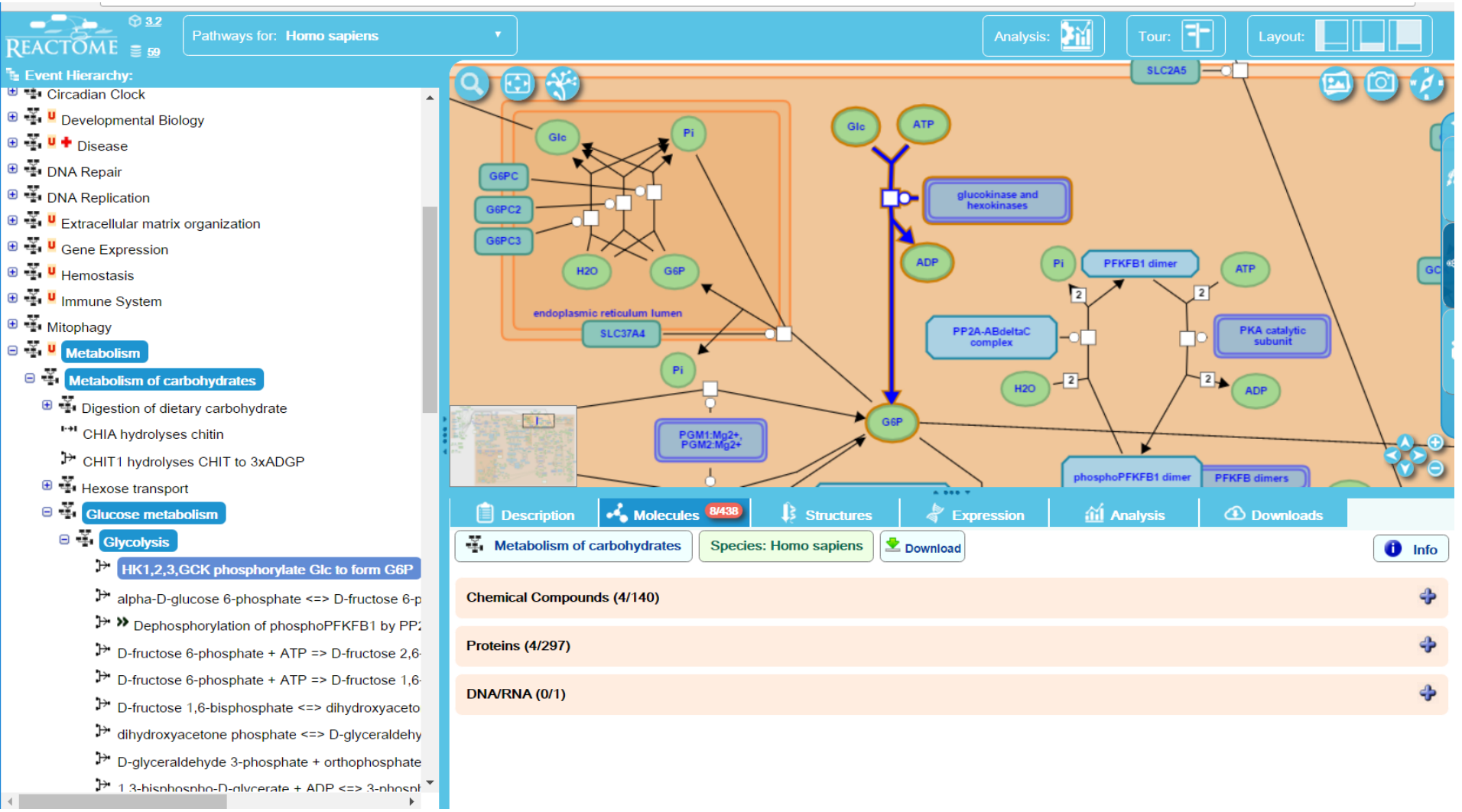
Иерархический список разделов (диаграмм)



Условные обозначения

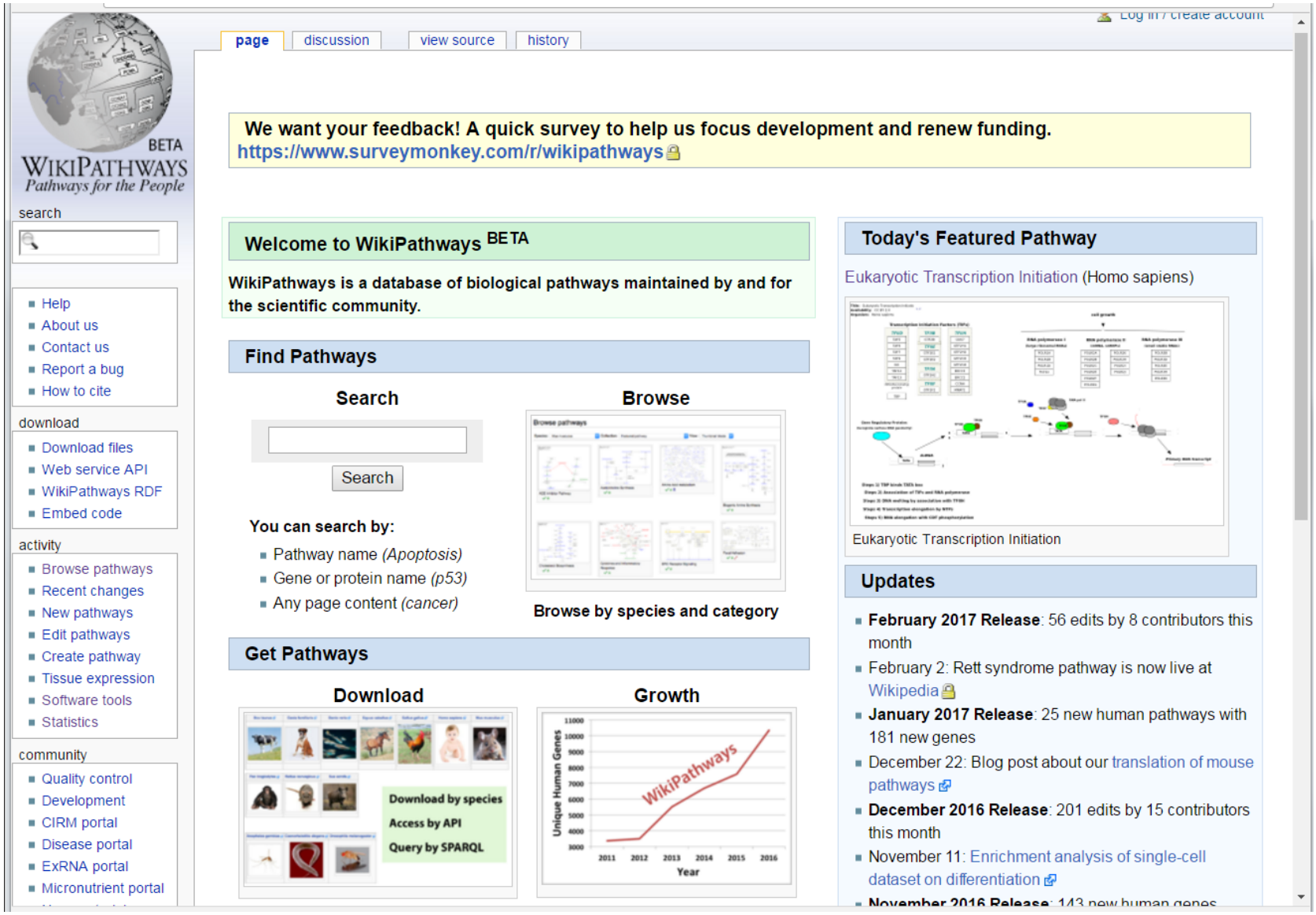
БАЗА ДАННЫХ REACTOME:

фрагмент диаграммы, представляющей реакции гликолиза



WikiPathways

http://www.wikipathways.org/index.php/WikiPathways



Статистика базы WikiPathways

Summary statistics

Last update: 2016/01/31

Number of pathways

- Public pathways: **2471**
- Private pathways: **4**

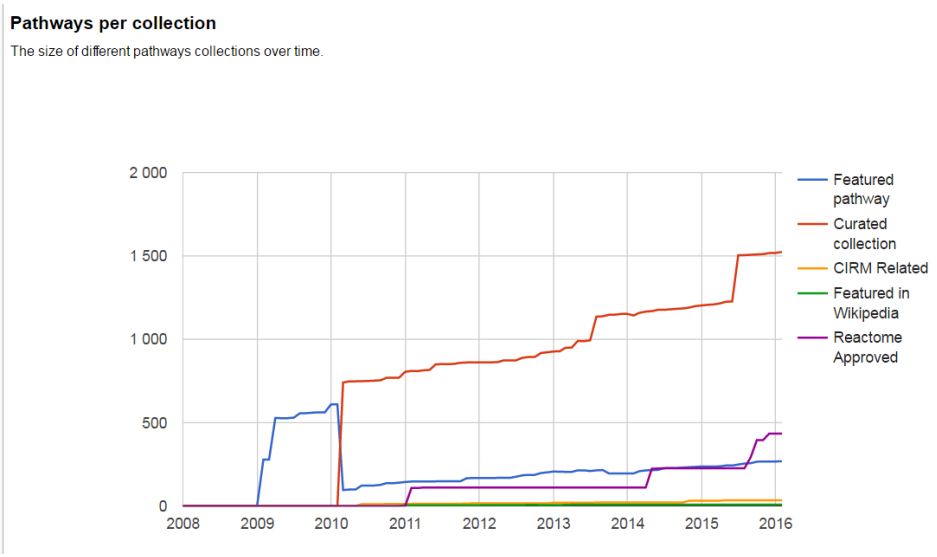
Number of active users

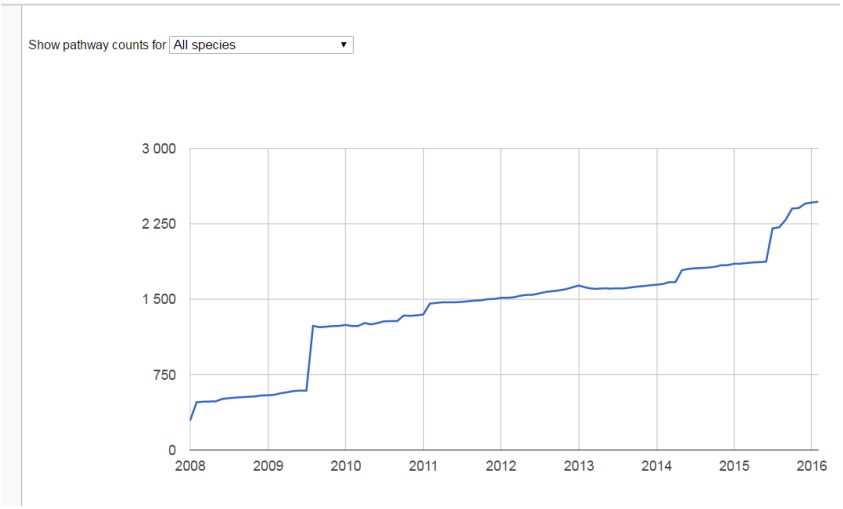
- At least 1 edit: **618**
- At least 1 edit (excluding test/tutorial pathways): **455**

Number of edits

- User edits: **21309**
- User edits (excluding test/tutorial pathways): **19529**
- Bot edits: **11418**

Неуклонный рост базы по годам

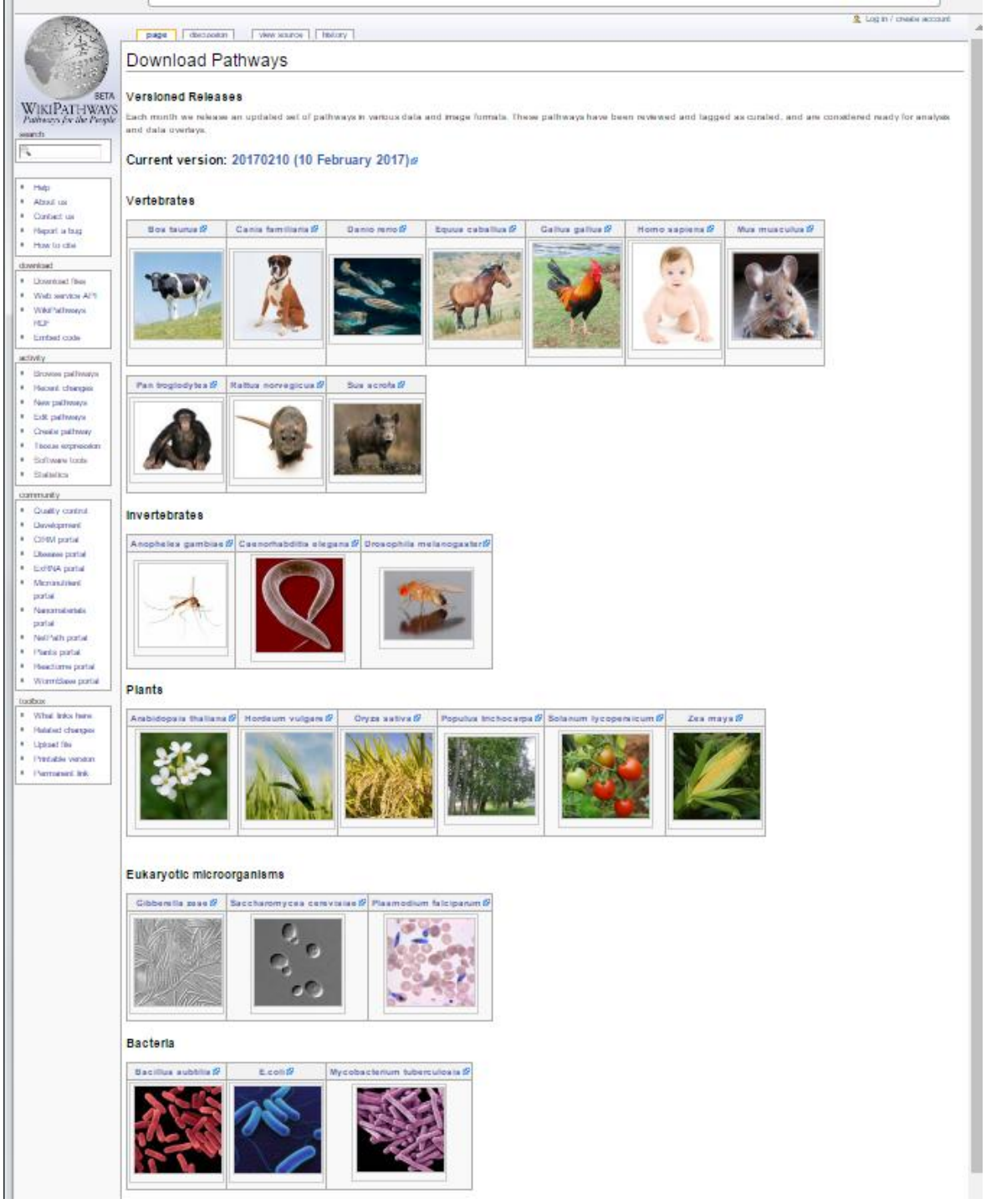




Kutmon M, Riutta A, Nunes N, Hanspers K, Willighagen EL, Bohler A, Mélius J, Waagmeester A, Sinha SR, Miller R, Coort SL, Cirillo E, Smeets B, Evelo CT, Pico AR. **WikiPathways: capturing the full diversity of pathway knowledge** *Nucl. Acids Res.*, 44, D488-D494 (**2016**)

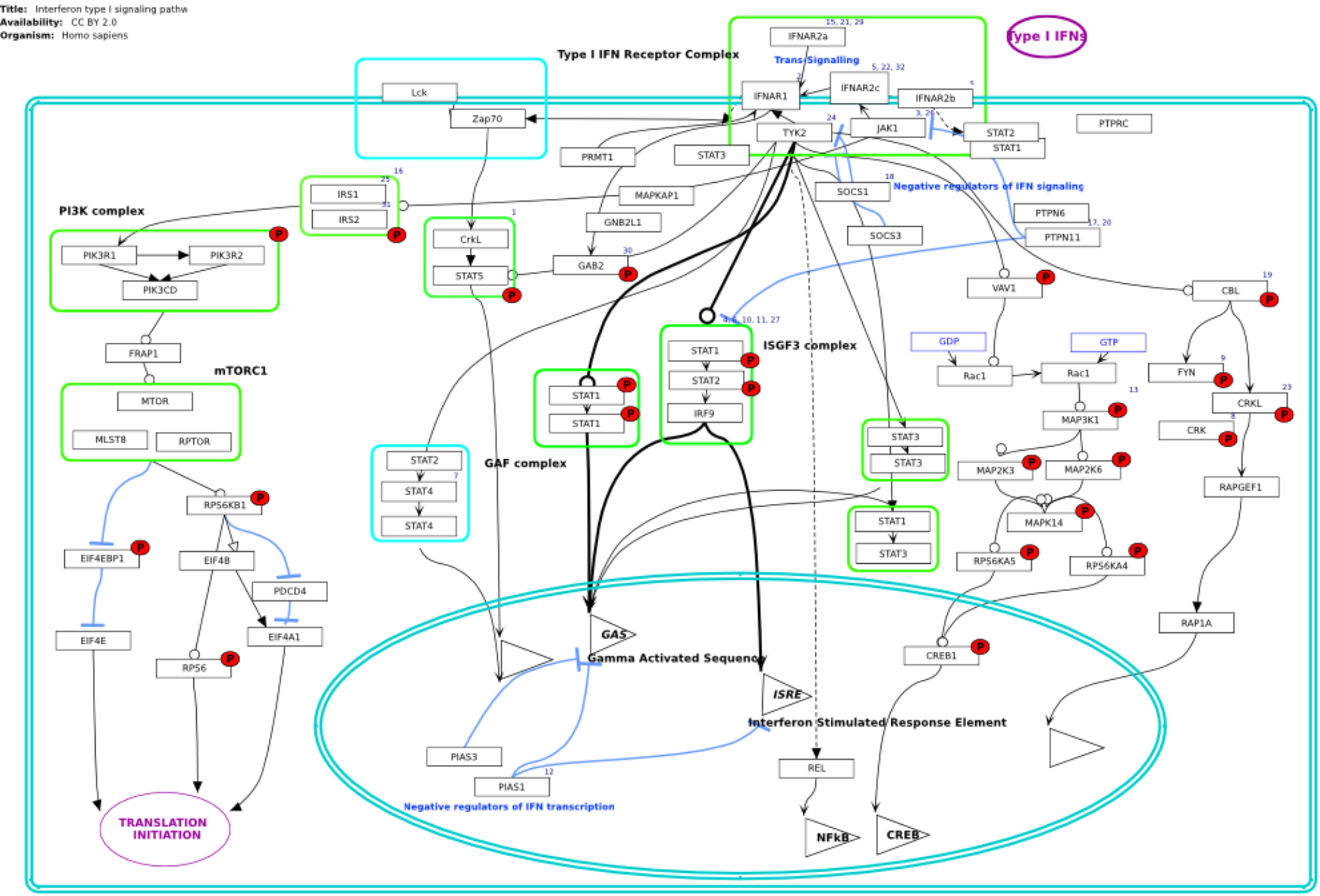
Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. **(2008)** WikiPathways: Pathway Editing for the People. PLoS Biol 6(7)

25 видов организмов базы WikiPathways



Пример представления данных в базе WikiPathways:

Диаграмма «Interferon type I signaling pathways (Homo sapiens)»



Все данные диаграмм WikiPathways можно скачать !!!!!

(Первый способ)

Выбор вида организма

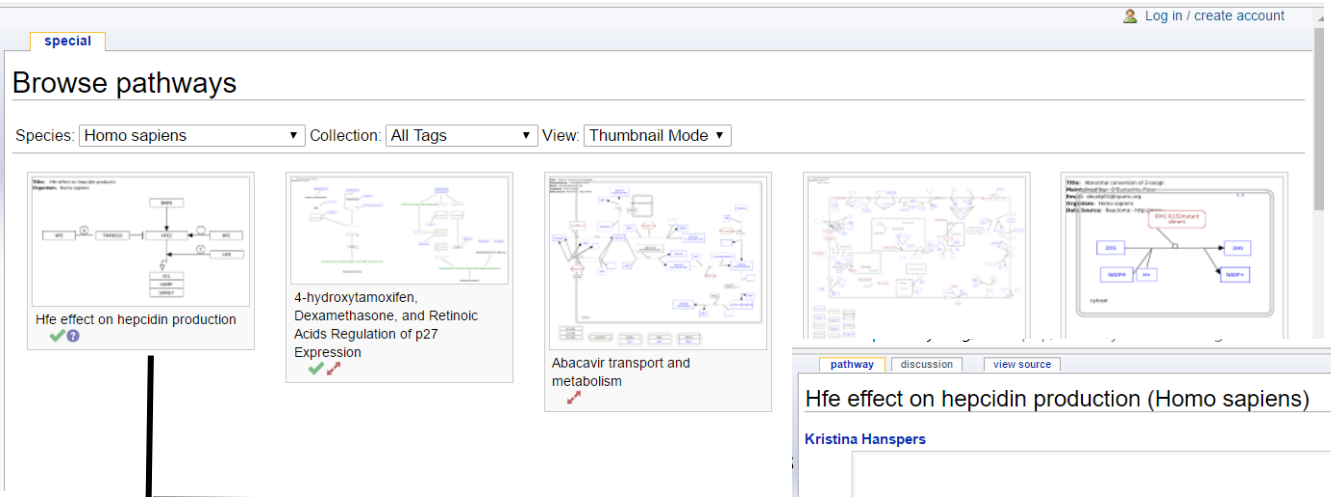
Архивный файл всех диаграмм одного вида организма (gpml формат)

Имя	Размер	Сжат	Тип	Изменён	CRC32
..			Папка с файл...		
Hs_4-hydroxytamoxifen,_Dexamethasone,_and_Retinoic_Acids_Regulation_of...	26 521	3 845	Файл "GPML"	13.10.2016 ...	BF8AD...
Hs_ACE_Inhibitor_Pathway_WP554_84372.gpml	37 697	5 343	Файл "GPML"	17.02.2016 ...	0728A...
Hs_Acetylcholine_Synthesis_WP528_79855.gpml	17 076	3 113	Файл "GPML"	04.05.2015 ...	1A63E...
Hs_Adipogenesis_WP236_80209.gpml	68 570	8 125	Файл "GPML"	22.05.2015 ...	B4C92...
Hs_Aflatoxin_B1_metabolism_WP699_70509.gpml	31 383	4 309	Файл "GPML"	04.05.2015 ...	DF49A...
Hs_AGE-RAGE_pathway_WP2324_89798.gpml	143 216	23 741	Файл "GPML"	06.10.2016 ...	6AAA3...
Hs_Alanine_and_aspartate_metabolism_WP106_91240.gpml	67 203	8 879	Файл "GPML"	21.01.2017 ...	85824...
Hs_Allograft_Rejection_WP2328_90020.gpml	160 876	20 090	Файл "GPML"	08.10.2016 ...	C9F45C...
Hs_Alpha_6_Beta_4_signaling_pathway_WP244_85199.gpml	74 687	9 742	Файл "GPML"	25.04.2016 ...	64D85...
Hs_Alzheimers_Disease_WP2059_87372.gpml	144 549	16 630	Файл "GPML"	22.07.2016 ...	C66543...
Hs_Amino_acid_conjugation_of_benzoic_acid_WP521_88588.gpml	9 727	1 726	Файл "GPML"	11.08.2016 ...	CF347B...
Hs_Amino_acid_conjugation_WP715_63154.gpml	5 556	1 132	Файл "GPML"	04.05.2015 ...	DDBB9...
Hs_Amino_Acid_metabolism_WP3925_90737.gpml	201 780	25 954	Файл "GPML"	13.12.2016 ...	32E4C9...
Hs_AMP-activated_Protein_Kinase_(AMPK)_Signaling_WP1403_90259.gpml	79 507	12 713	Файл "GPML"	27.10.2016 ...	73E441...
Hs_Amplification_and_Expansion_of_Oncogenic_Pathways_as_Metastatic_Trait...	22 371	4 143	Файл "GPML"	13.10.2016 ...	947498...
Hs_Amyotrophic_lateral_sclerosis_(ALS)_WP2447_85186.gpml	63 458	10 080	Файл "GPML"	23.04.2016 ...	314524...
Hs_Androgen_receptor_signaling_pathway_WP138_79958.gpml	105 471	16 748	Файл "GPML"	04.05.2015 ...	B374D...
Hs_Angiogenesis_WP1539_88983.gpml	47 789	6 560	Файл "GPML"	18.08.2016 ...	2A716...
Hs_Angiopoietin_Like_Protein_8_Regulatory_Pathway_WP3915_90629.gpml	183 173	25 164	Файл "GPML"	02.12.2016 ...	BBDDF...
Hs_ApoE_and_miR-146_in_inflammation_and_atherosclerosis_WP3926_90739....	16 810	3 236	Файл "GPML"	13.12.2016 ...	F98AC...
Hs_Apoptosis_Modulation_and_Signaling_WP1772_91293.gpml	90 928	9 952	Файл "GPML"	23.01.2017 ...	9ED92...
Hs_Apoptosis_Modulation_by_HSP70_WP384_67054.gpml	34 491	4 874	Файл "GPML"	27.04.2015 ...	B15FE...
Hs_Apoptosis_WP254_88977.gpml	74 141	9 805	Файл "GPML"	18.08.2016 ...	A06A1...
Hs_Apoptosis-related_network_due_to_altered_Notch3_in_ovarian_cancer_WP...	46 898	5 207	Файл "GPML"	04.05.2015 ...	EB991E...
Hs_Arachidonate_Epoxygenase_-_Epoxide_Hydrolase_WP678_71506.gpml	38 335	5 169	Файл "GPML"	04.05.2015 ...	2C7AD...
Hs_Arrhythmogenic_Right_Ventricular_Cardiomyopathy_WP2118_71265.gpml	97 314	9 064	Файл "GPML"	04.05.2015 ...	C856B...
Hs_Aryl_Hydrocarbon_Receptor_Pathway_WP2873_88902.gpml	92 983	10 444	Файл "GPML"	17.08.2016 ...	5AF732...
Hs_Aryl_Hydrocarbon_Receptor_WP2586_89793.gpml	158 322	19 370	Файл "GPML"	06.10.2016 ...	F4D3EF...
Hs_Arylamine_metabolism_WP694_89536.gpml	18 479	2 823	Файл "GPML"	16.09.2016 ...	4EA6D...
Hs_Association_Between_Physico-Chemical_Features_and_Toxicity_Associated...	62 386	8 420	Файл "GPML"	13.10.2016 ...	D6733...
Hs_ATM_Signaling_Network_in_Development_and_Disease_WP3878_89745.g...	42 186	5 966	Файл "GPML"	29.09.2016 ...	B48E87...
Hs_ATM_Signaling_Pathway_WP2516_90247.gpml	295 343	34 278	Файл "GPML"	26.10.2016 ...	0FAF71...

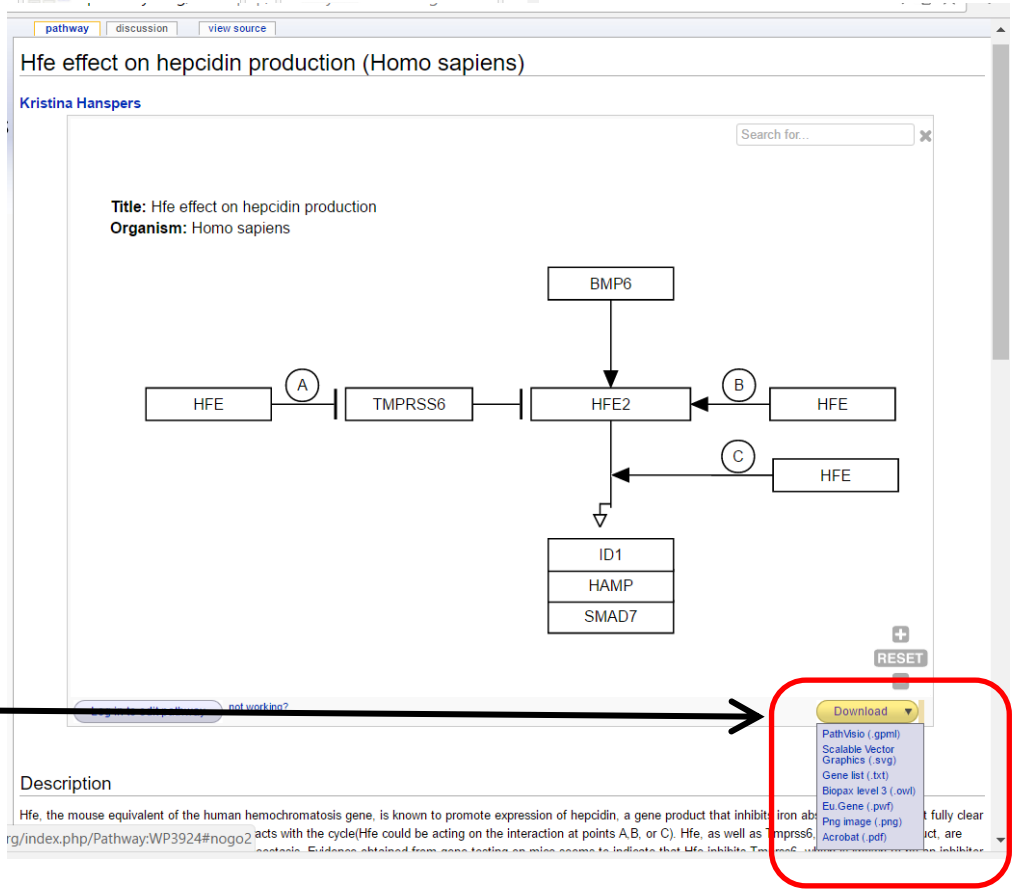
Всего: 25 661 938 байт в 379 файлах

Данные диаграмм WikiPathways можно все скачать !!!!!

(Второй способ)

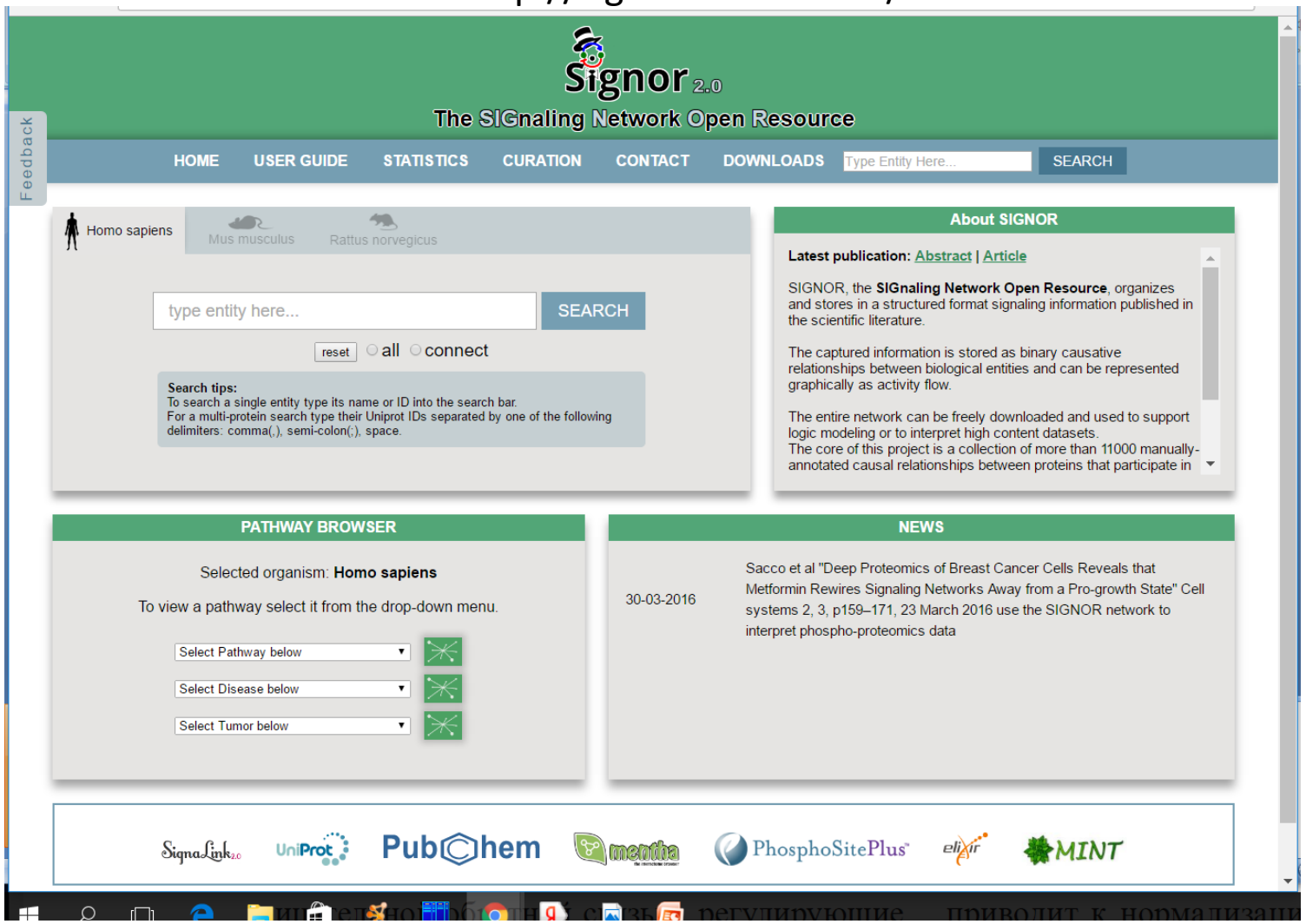


Скачиваем каждую диаграмму поотдельности в форматах:
gpml,
svg,
txt,
owl,
pwf,
png,
pdf



РЕСУРС Signor

http://signor.uniroma2.it/

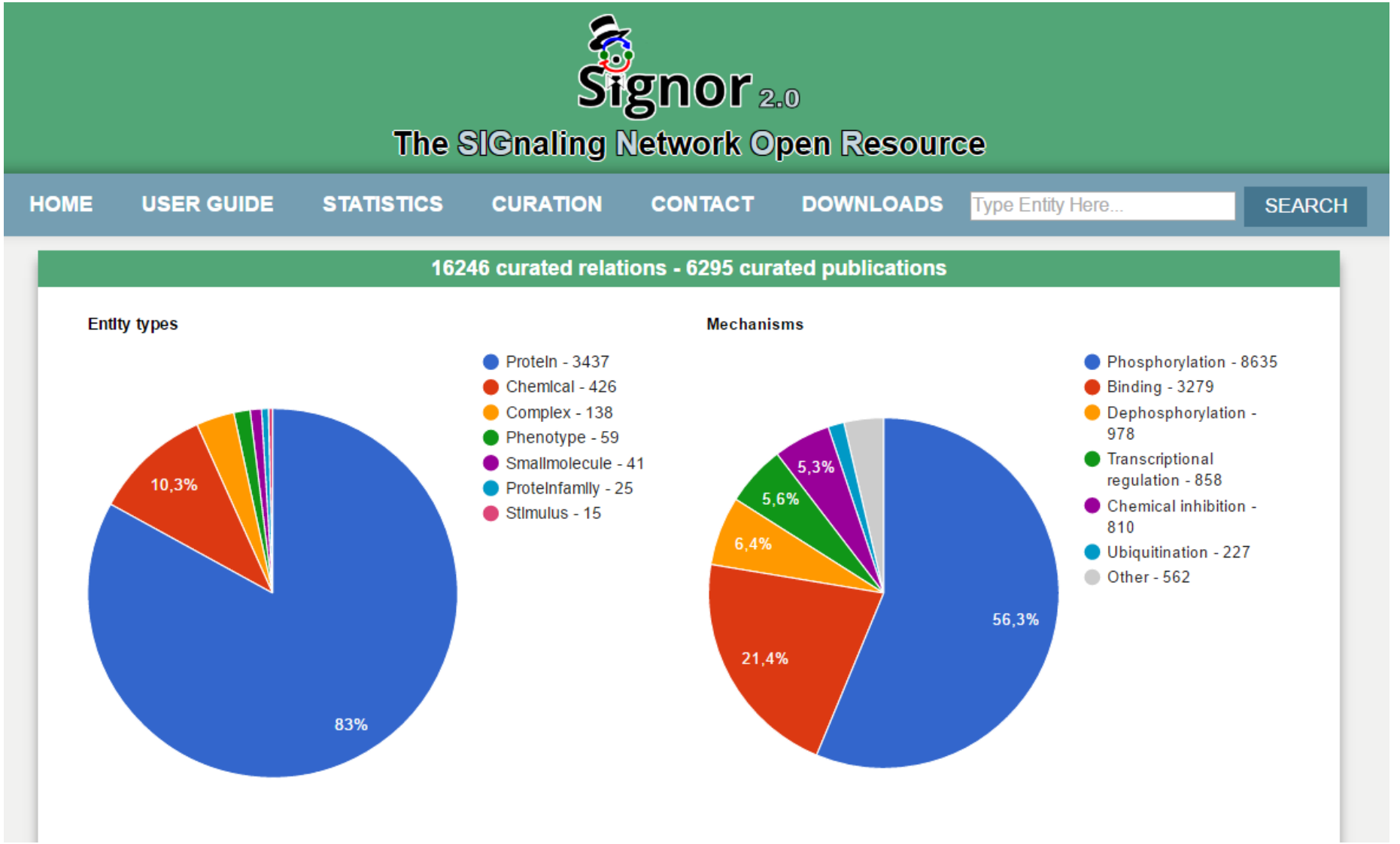


База данных по сигнальным путям трех видов организмов (человек , мышь, крыса). Данные взяты из научных публикаций

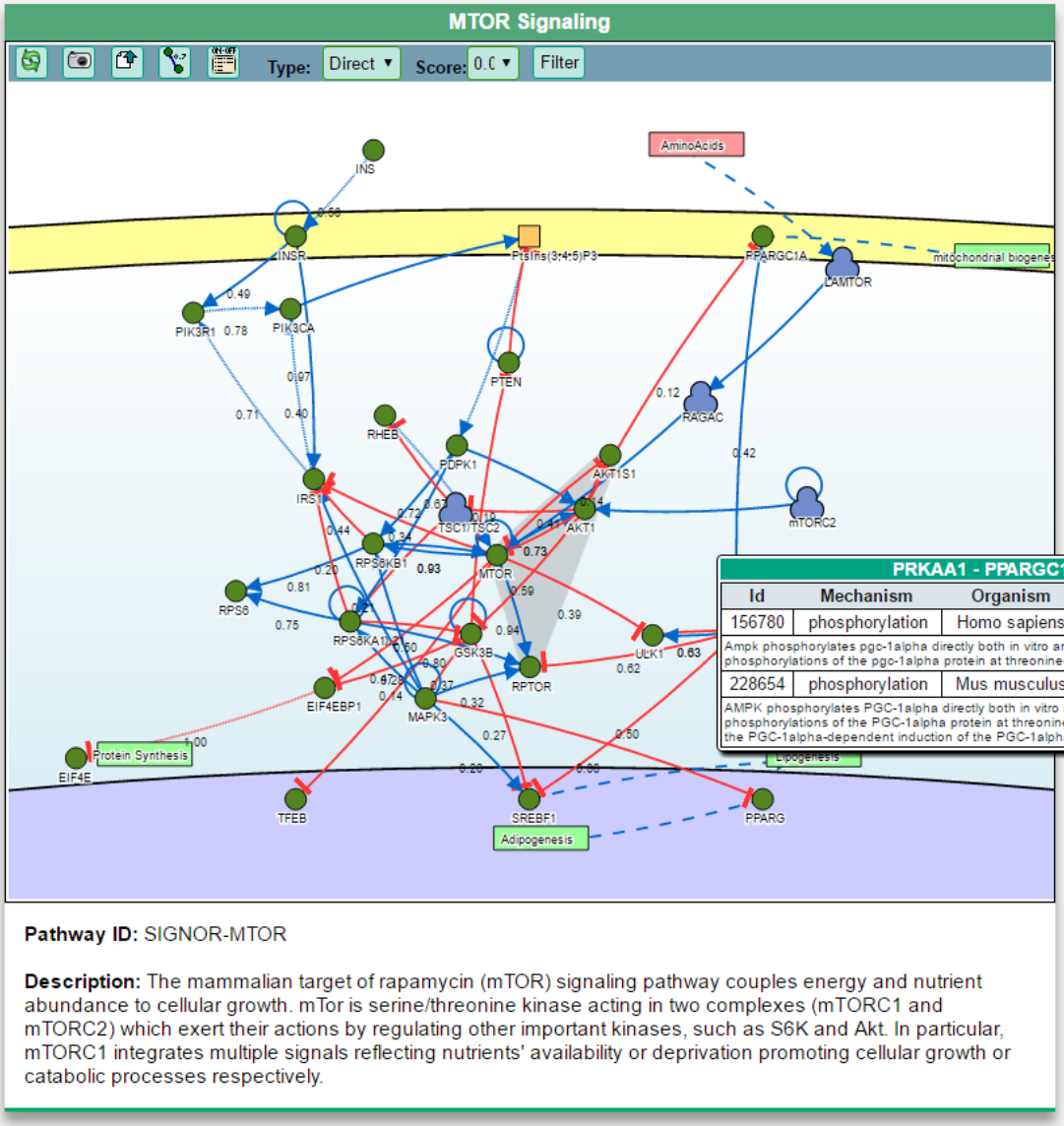
Perfetto L et al., SIGNOR: a database of causal relationships between biological entities. Nucleic Acids Res. **2016**;44(D1):D548-54.

Статистика данных в базе Signor

http://signor.uniroma2.it/statistics.php



Ресурс Signor : представление данных по mTOR сигнальному пути



Pathway ID: SIGNOR-MTOR

Description: The mammalian target of rapamycin (mTOR) signaling pathway couples energy and nutrient abundance to cellular growth. mTor is serine/threonine kinase acting in two complexes (mTORC1 and mTORC2) which exert their actions by regulating other important kinases, such as S6K and Akt. In particular, mTORC1 integrates multiple signals reflecting nutrients' availability or deprivation promoting cellular growth or catabolic processes respectively.

37 Seed Entities

Name	Primary ID
Organism:	
Add Entity	
Adipogenesis	SIGNOR-PH26
AKT1	P31749
AKT1S1	Q96B36
AminoAcids	SIGNOR-ST5
EIF4E	P06730
EIF4EBP1	Q13541
GSK3B	P49841
HIF1AN	Q9NWT6
INS	P01308
	P06213
	P35568
	SIGNOR-C26
	SIGNOR-PH30
	P27361
mitochondrial biogenesis	SIGNOR-PH32
MTOR	P42345
mTORC2	SIGNOR-C2
PDPK1	O15530
PIK3CA	P42336
PIK3R1	P27986
PPARG	P37231
PPARGC1A	Q9UBK2
PRKAA1	Q13131
PRKAG1	P54619

Ресурс Signor : представление данных о путях активации фактора SREBP1

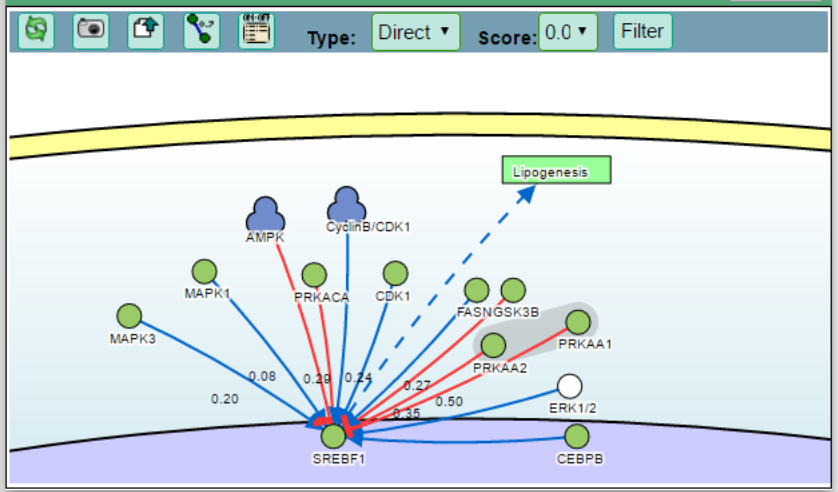
Signor 2.0 The SIGnaling Network Open Resource

HOME | USER GUIDE | STATISTICS | CURATION | CONTACT | DOWNLOADS | Type Entity Here... | SEARCH

Summary

Name	SREBF1 — View Modifications
Full Name	Sterol regulatory element-binding protein 1
Synonyms	SREBP-1, Class D basic helix-loop-helix protein 1, bHLHd1, Sterol regulatory element-binding transcription factor 1 | BHLHD1, SREBP1
Primary ID	P36956
Links	UniProt - GeneCards - mentha
Type	protein
Relations	26

Viewer Expand



Search Tools

Refine search:

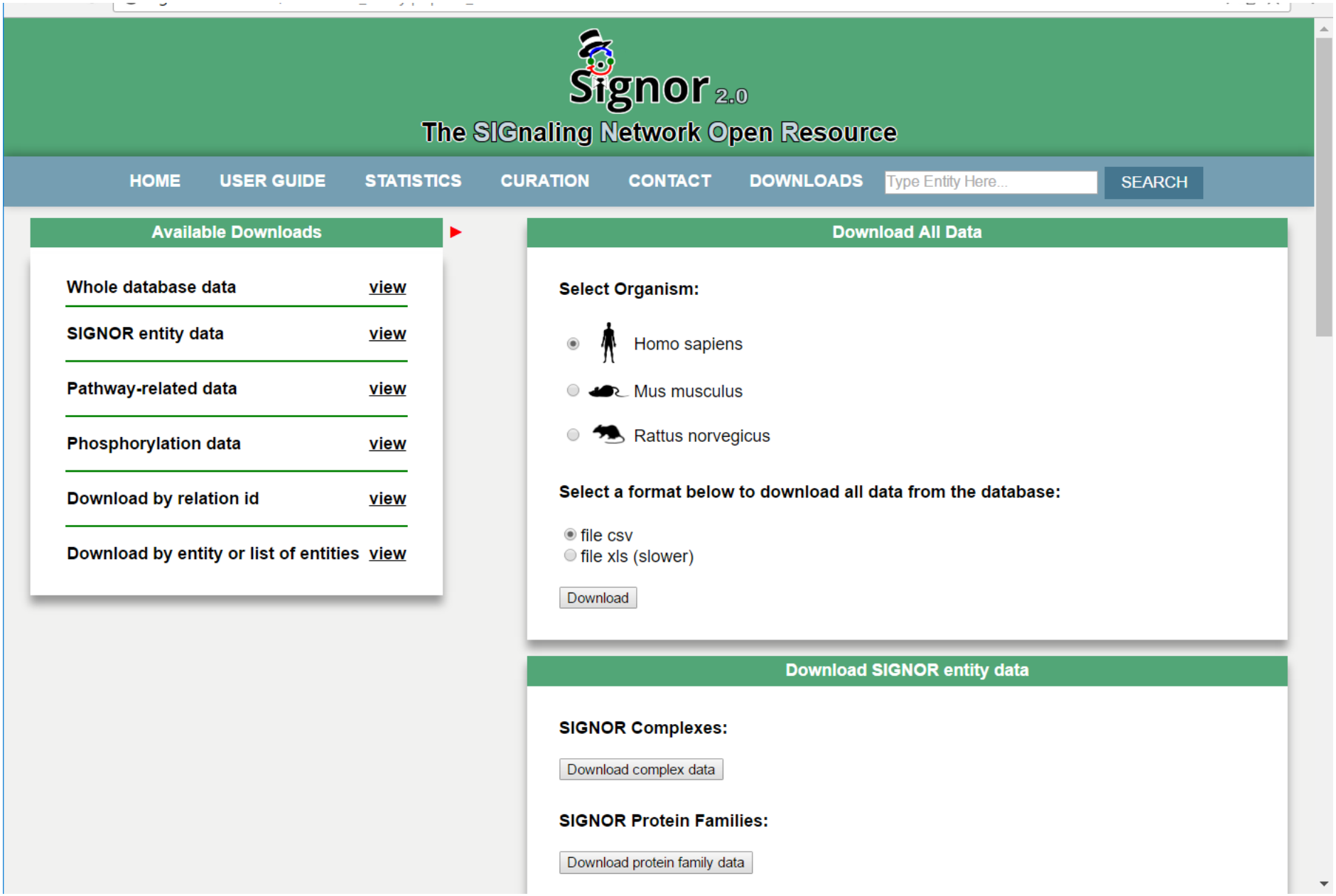
- by mechanism: none
- by effect: none
- by organism: none

Modifications Tables

Relations

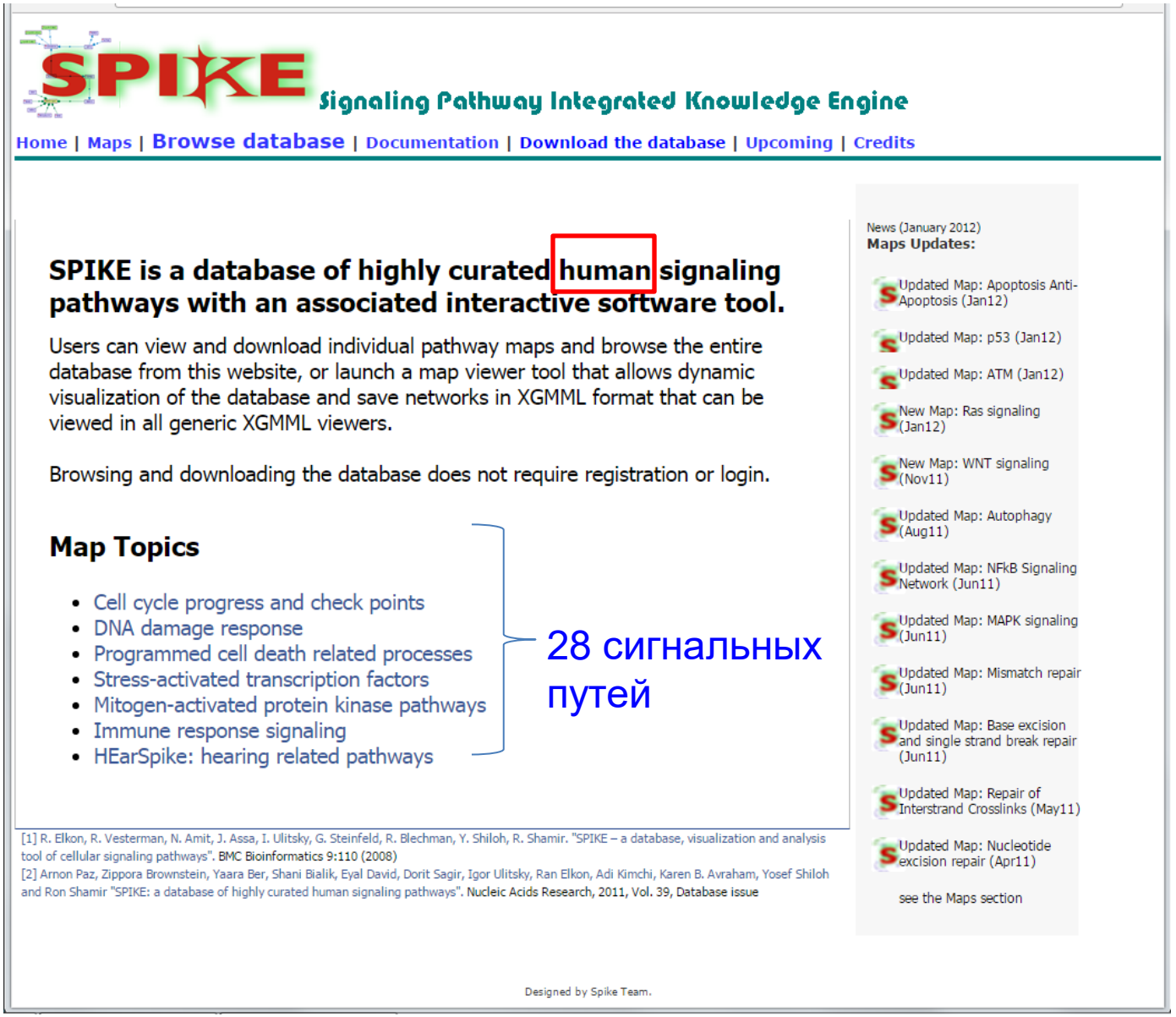
	REGULATOR	MECHANISM	TARGET	SCORE
+	MAPK3	up-regulates activity → phosphorylation	SREBF1	0.20
Publications: 1			Organism:	Homo Sapiens
Pathways:	MTOR Signaling			
+	MAPK1	up-regulates → phosphorylation	SREBF1	0.08
Publications: 1			Organism:	Homo Sapiens

Ресурс Signor : данные можно скачать



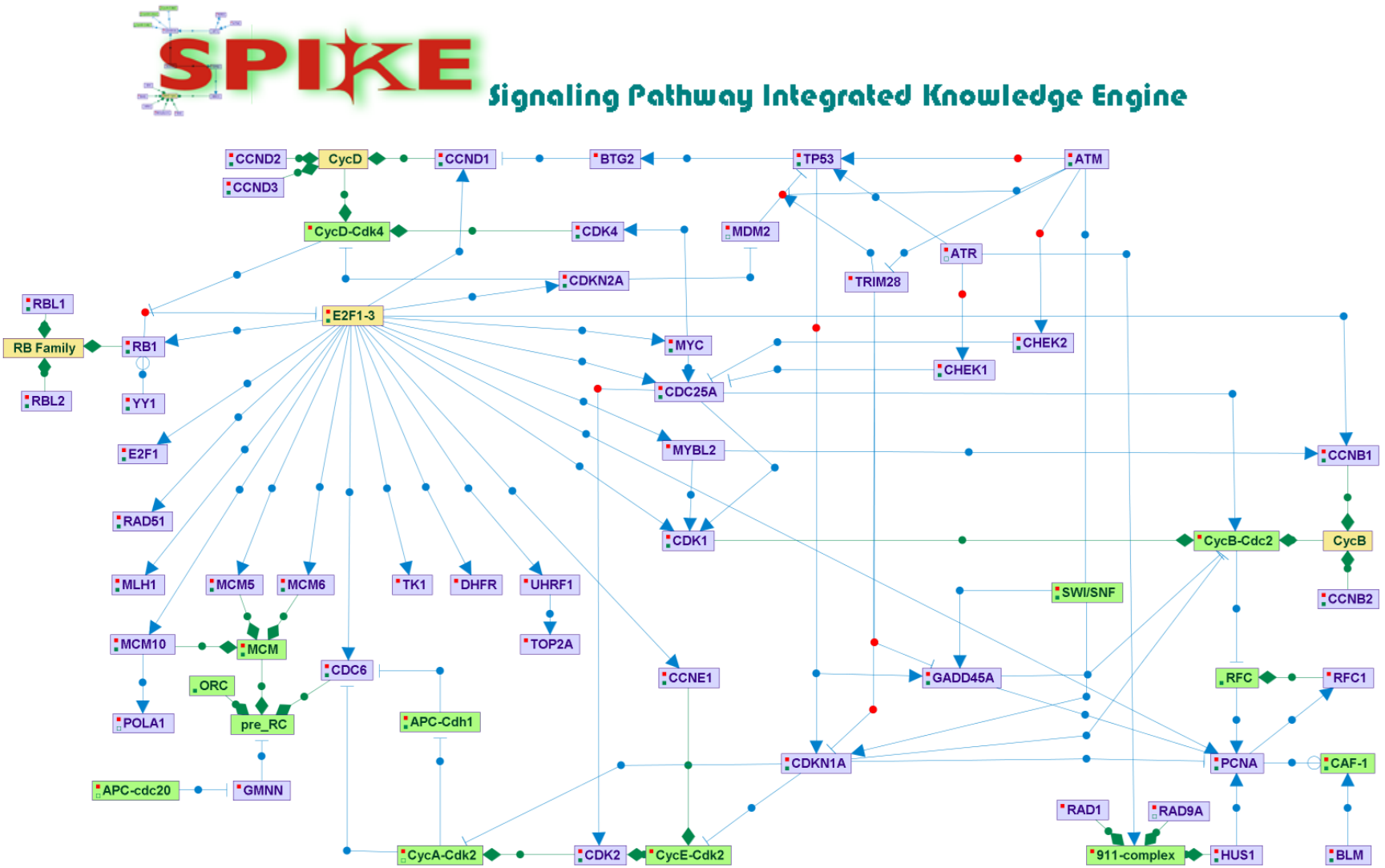
РЕСУРС SPIKE

http://www.cs.tau.ac.il/~spike/

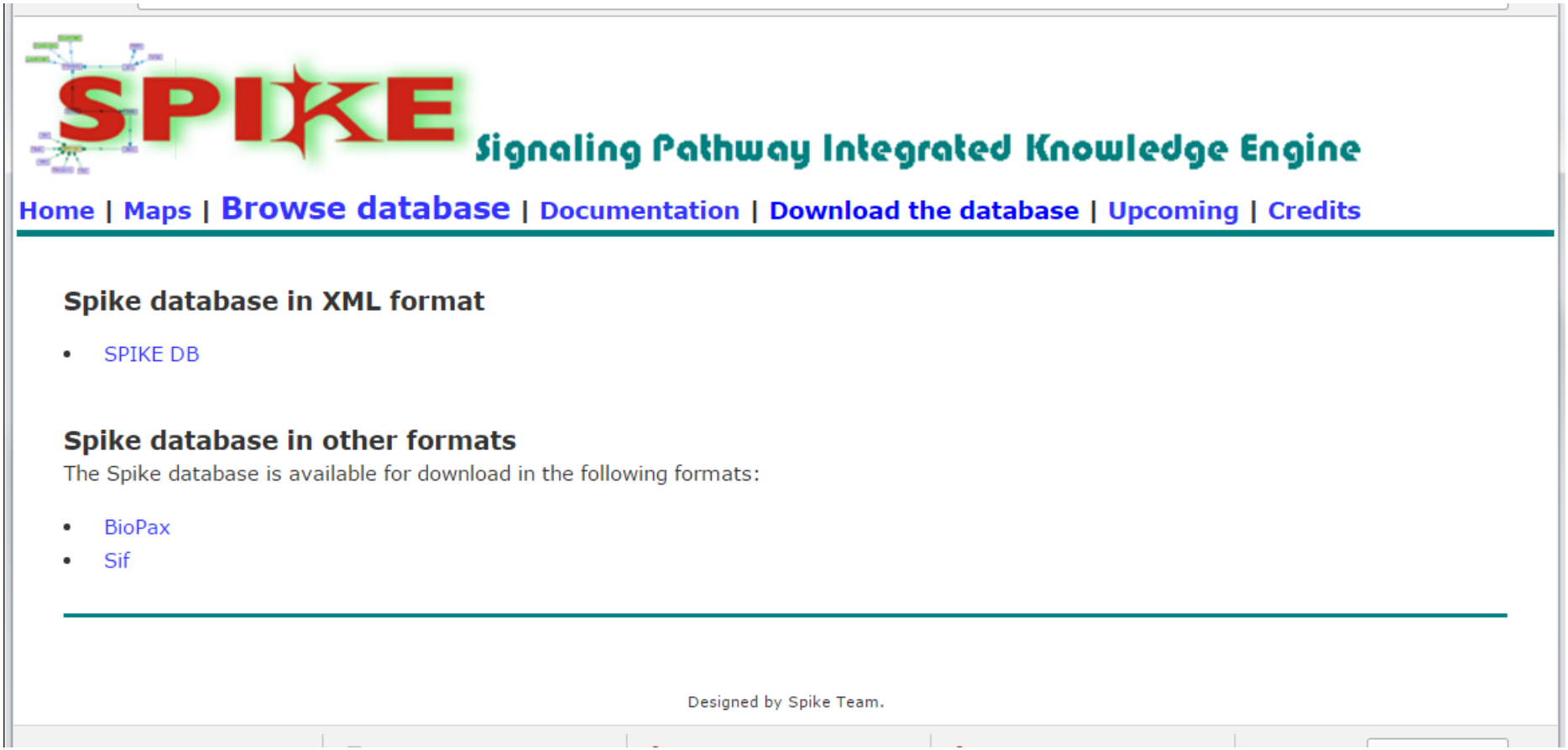


Paz A. et al., SPIKE: a database of highly curated human signaling pathways Nucleic Acids Research, **2011**, Vol. 39, Database issue

База SPIKE: Представление данных о процессе «DNA damage induced G1-S checkpoint (spike00003) »

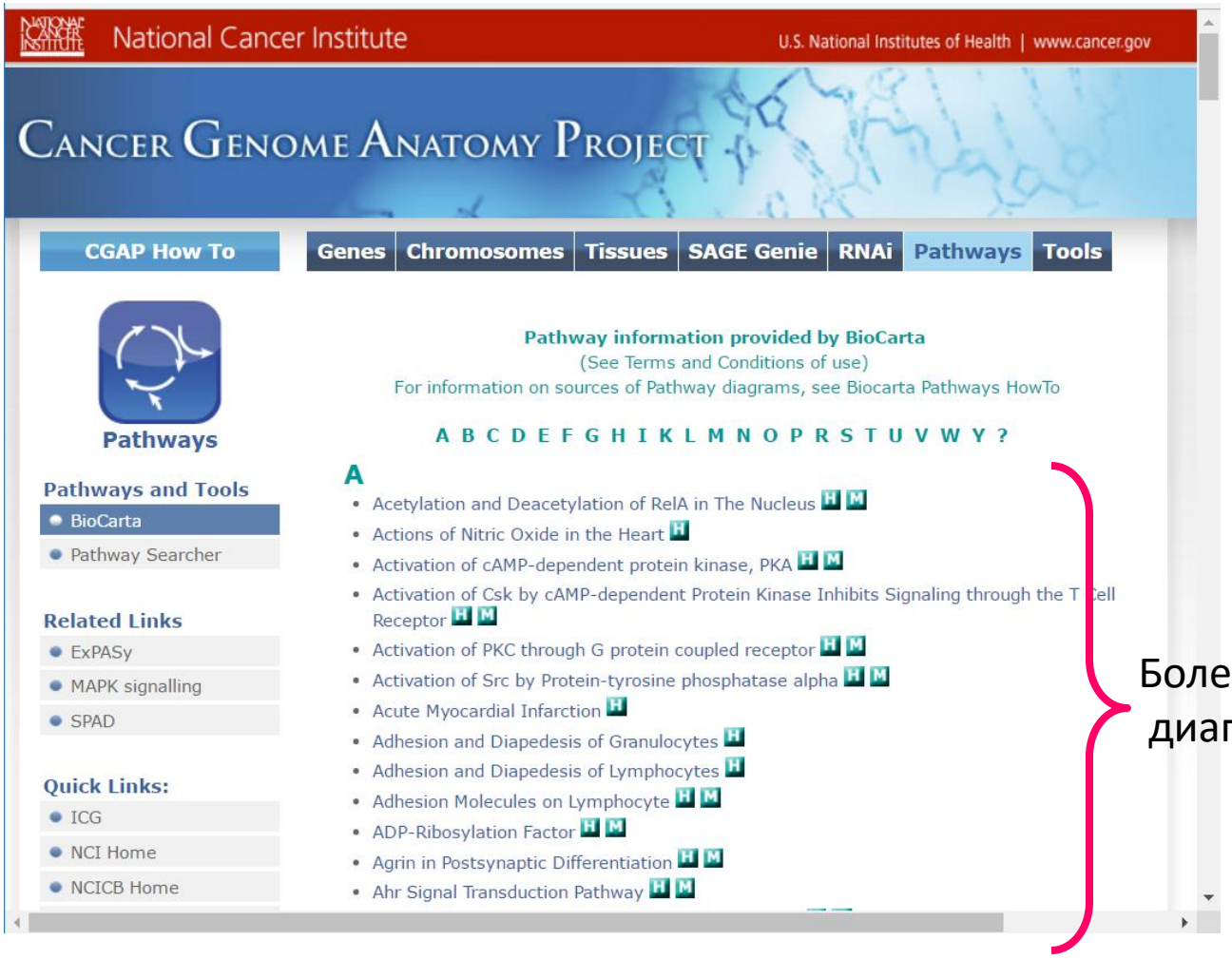


Ресурс SPIKE : данные можно скачать



ИНФОРМАЦИОННЫЙ РЕСУРС BIOCARTA

https://cgap.nci.nih.gov/Pathways/BioCarta_Pathways



Более 300 диаграмм

БАЗА НЕ ОБНОВЛЯЕТСЯ:

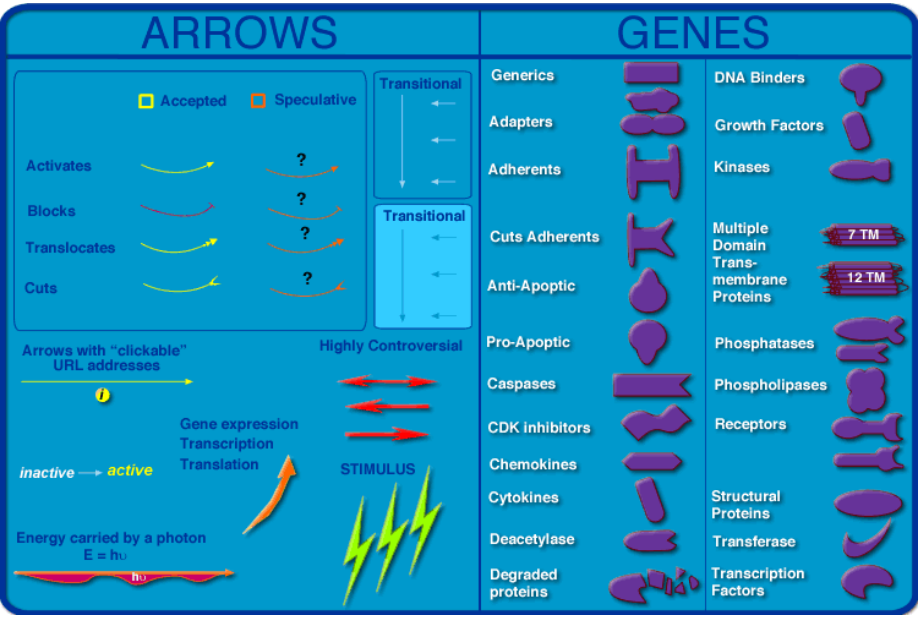
BioCarta had not been updating its pathways. The information provided might have been outdated. As a result, we have discontinued offering pathway information online. You may view our pathway figures at http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways.

http://www.biocarta.com/

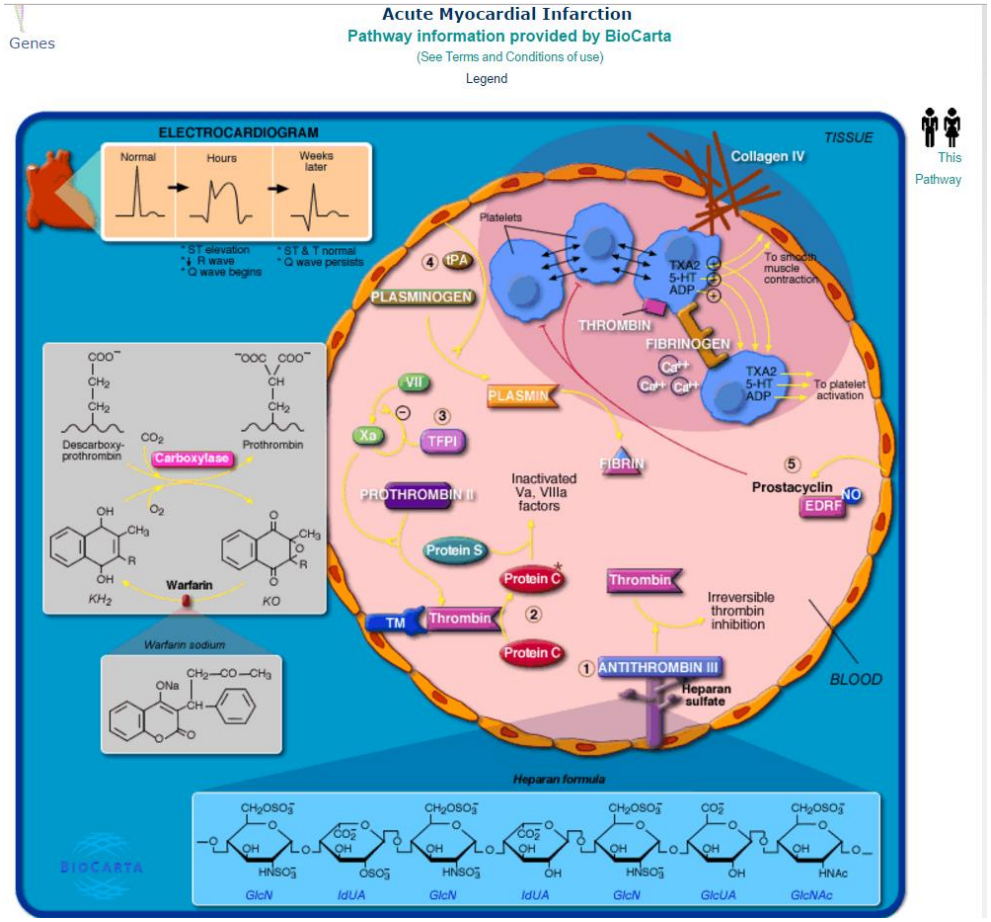
ИНФОРМАЦИОННЫЙ РЕСУРС BIOCARTA

Содержит схемы метаболических либо сигнальных путей, а также схемы процессов

Условные обозначения



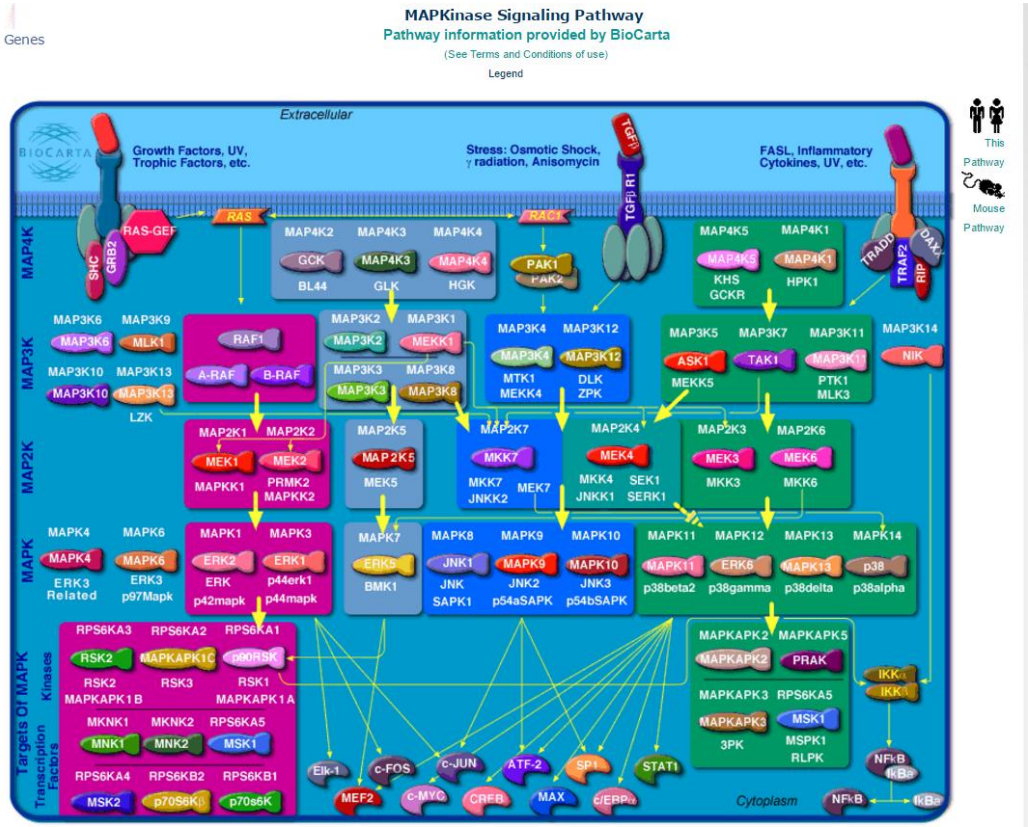
Acute Myocardial Infarction



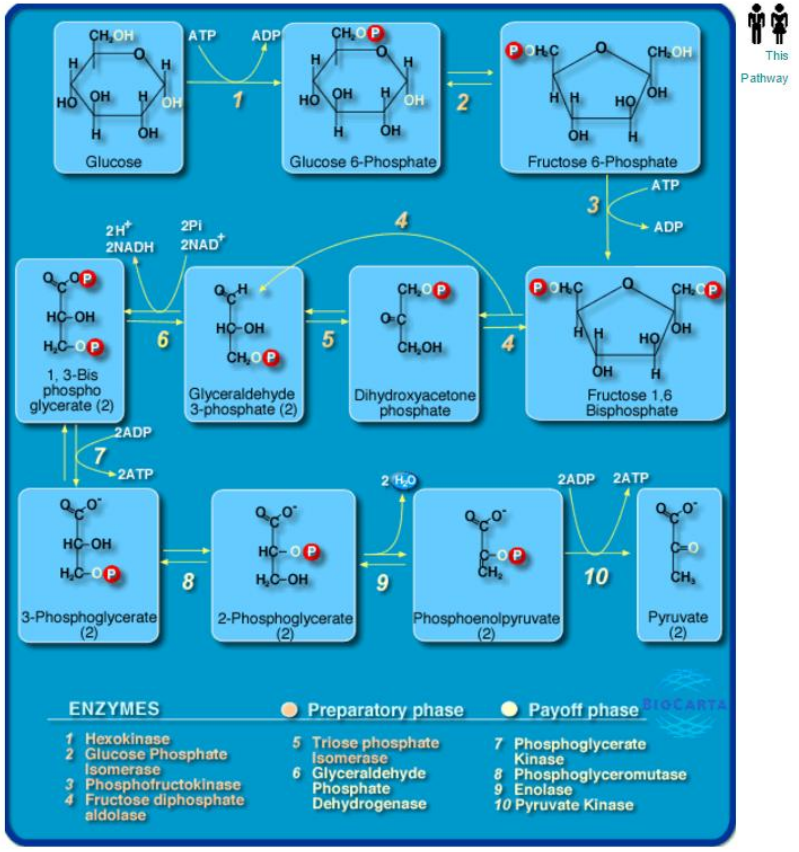
ИНФОРМАЦИОННЫЙ РЕСУРС BIOCARTA (продолжение)

Содержит схемы метаболических либо сигнальных путей, а также схемы процессов

MAPKinase Signaling Pathway



Glycolysis Pathway



ПОВТОРЕНИЕ: В лекции № 4 были рассмотрены базы данных по генным сетям, а также по метаболическим и сигнальным путям:

1) GeneNet – ИЦиГ СО РАН , г.Новосибирск
http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/

2) KEGG Kyoto encyclopedia of genes and genomes:
integrated suite of databases on genes, proteins, and metabolic pathways
http://www.genome.ad.jp/kegg

3) MetaCyc Metabolic Database http://metacyc.org/
+ BioCyc (Database Collection) https://biocyc.org/

4) Reactome http://www.reactome.org/

5) WikiPathways http://www.wikipathways.org/index.php/WikiPathways

7) Signor http://signor.uniroma2.it/

8) SPIKE http://www.cs.tau.ac.il/~spike/

9) BioCarta https://cgap.nci.nih.gov/Pathways/BioCarta_Pathways

СПАСИБО ЗА ВНИМАНИЕ !