

Орг. вопросы

- Сайт <http://intbio.org/bioinf2018-2019>
- **Вводный опросник** – дедлайн 12 октября 23:59.
- **Форма отчетности:** в этом семестре зачет

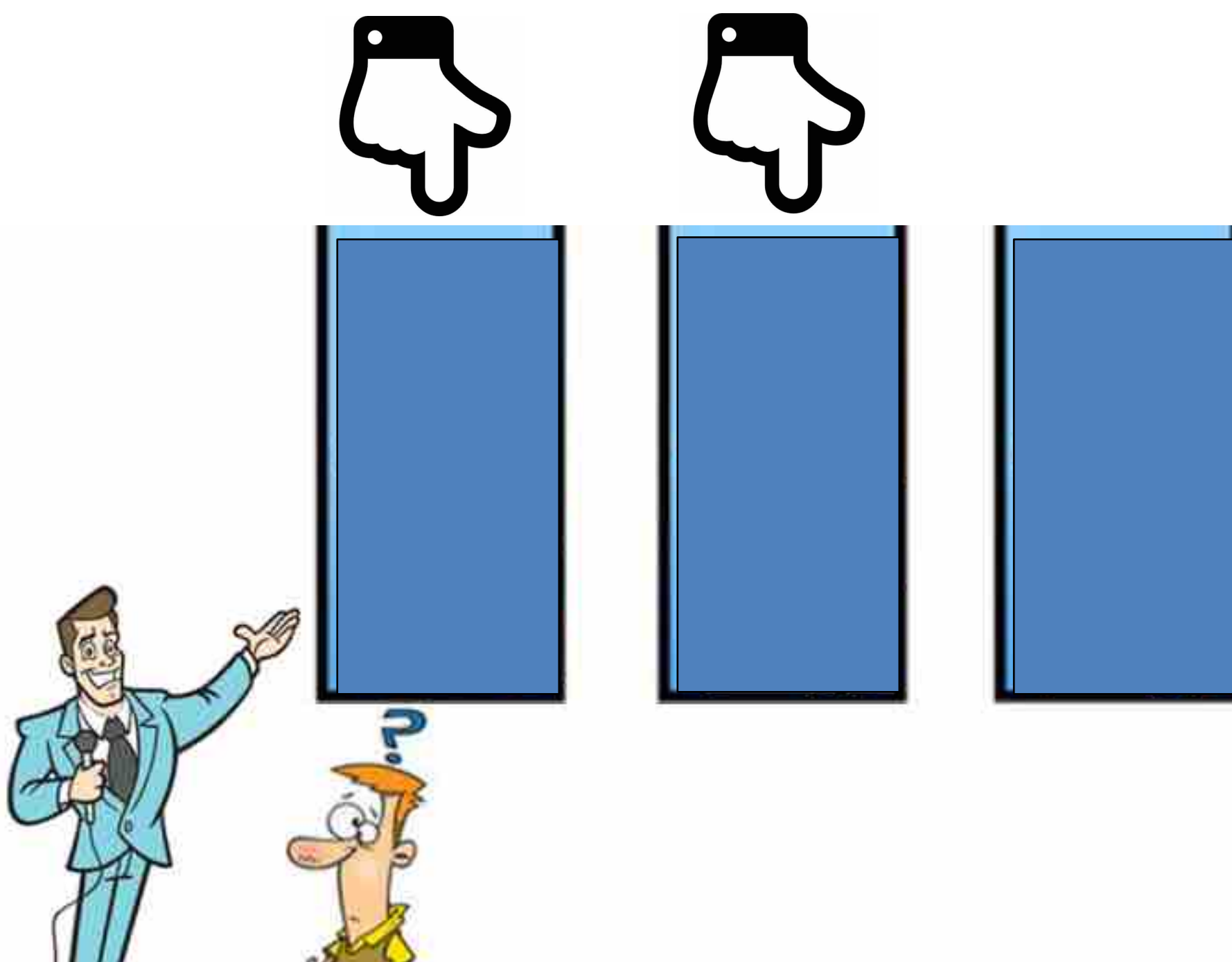
ВВЕДЕНИЕ В БИОИНФОРМАТИКУ

Лекции №2-3

Природа, передача и хранение информации. Базы данных. Биологические БД. Базы данных NCBI.

Алексей Константинович Шайтан, к.ф.-м.н.

Сайт курса: <http://intbio.org/bioinf2018-2019>



- В результате игрок получает **0.67 бита информации**
- Если бы ведущий открыл дверь в самом начале – только **0.58 бита информации**

Природа информации

Информация

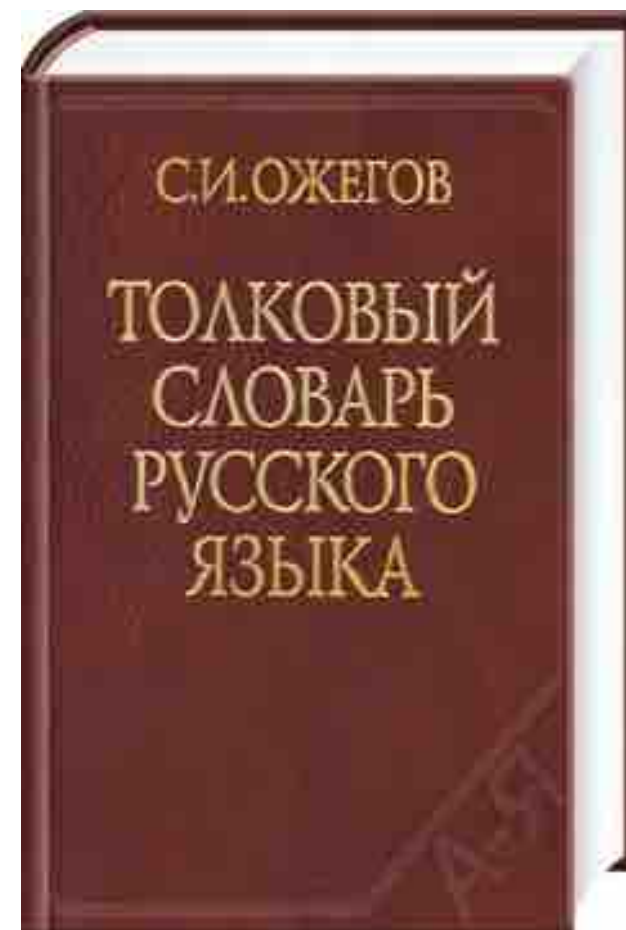
ИНФОРМАЦИЯ, 1. Сведения об окружающем мире и протекающих в нем процессах, воспринимаемые человеком или специальным устройством.



Н.Н. Моисеев

... универсального определения информации не только нет, но и быть не может из-за широты этого понятия.

Information is information, not matter or energy.



Norbert Wiener

Информация



iPhone 7 Plus

Серебристый

Ёмкость

Теперь выберите ёмкость.

32 ГБ¹

52 990.00 руб.

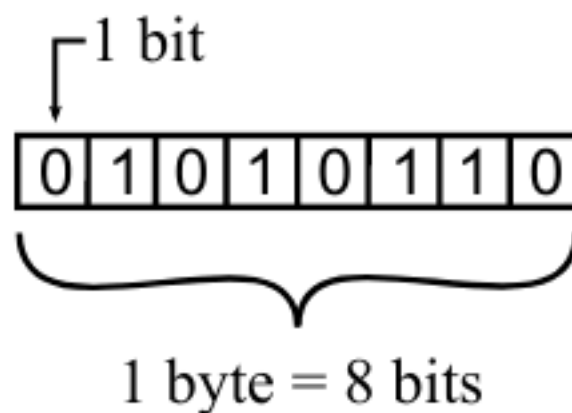
Доставка: на складе

128 ГБ¹

60 990.00 руб.

Доставка: на складе

Измерение информации



Измерения в байтах				
ГОСТ 8.417—2002			Приставки СИ	
Название	Обозначение	Степень	Название	Степень
байт	Б	10^0	-	10^0
килобайт	кбайт	10^3	кило-	10^3
мегабайт	Мбайт	10^6	мега-	10^6
гигабайт	Гбайт	10^9	гига-	10^9
терабайт	Тбайт	10^{12}	тера-	10^{12}
петабайт	Пбайт	10^{15}	пета-	10^{15}
эксабайт	Эбайт	10^{18}	экса-	10^{18}
зеттабайт	Збайт	10^{21}	зетта-	10^{21}
иоттабайт	Ибайт	10^{24}	иотта-	10^{24}

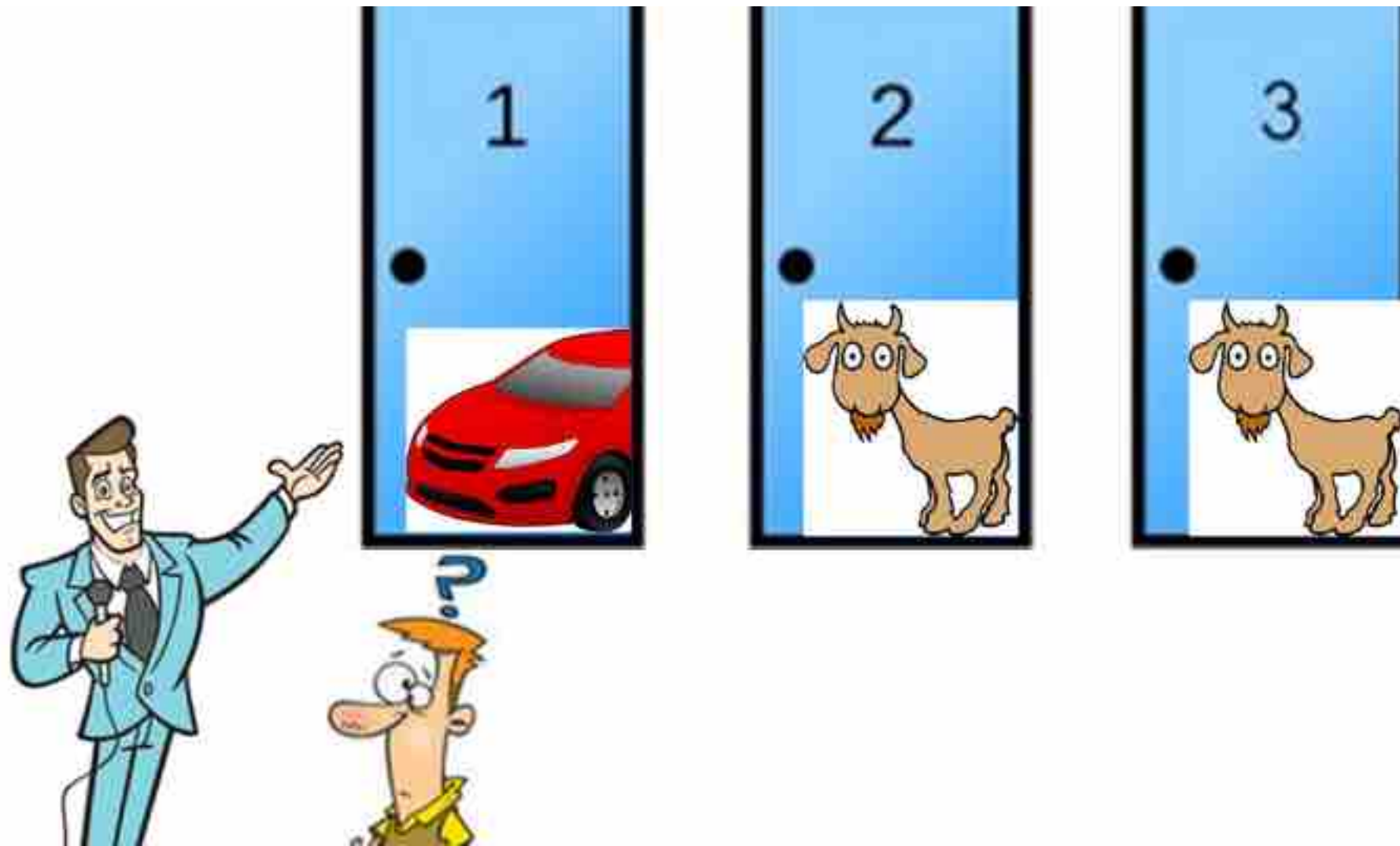
Decimal	Binary
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

Вероятность того,
что машина за
этой дверью:

$1/3$

$1/3$

$1/3$

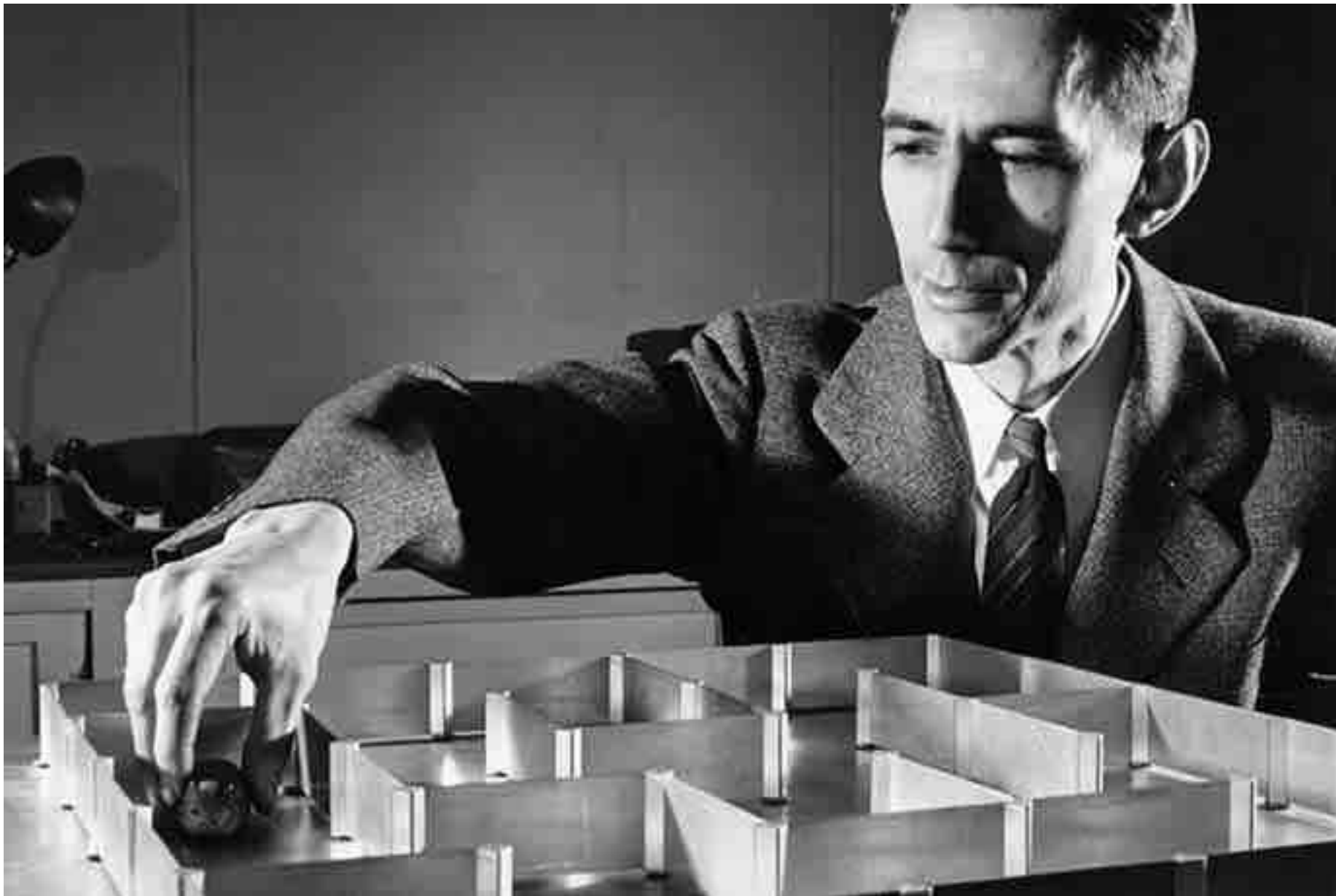


- Сколько информации нужно, чтобы закодировать
положении машины?

1.5849625007211563... бит

Теория информации

“the father of [information theory](#)”



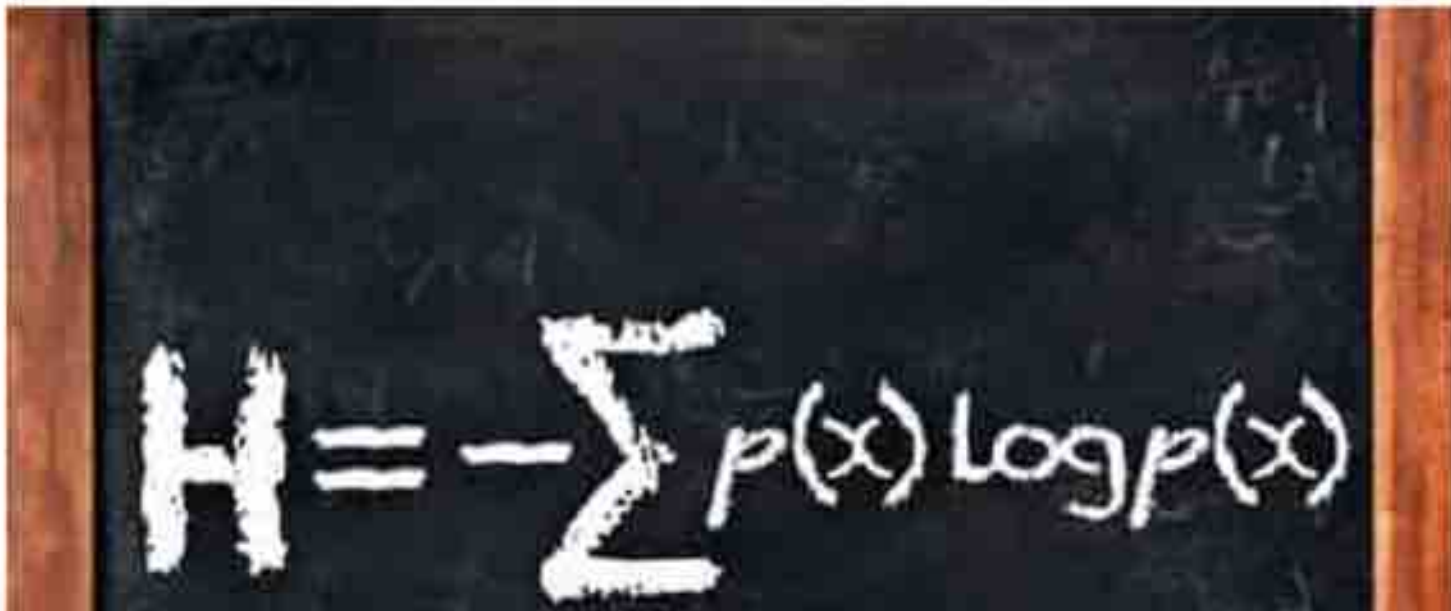
Claude Elwood Shannon
(April 30, 1916 – February 24, 2001)

Теория информации

Science A short history of equations

Without Claude Shannon's information theory there would have been no internet

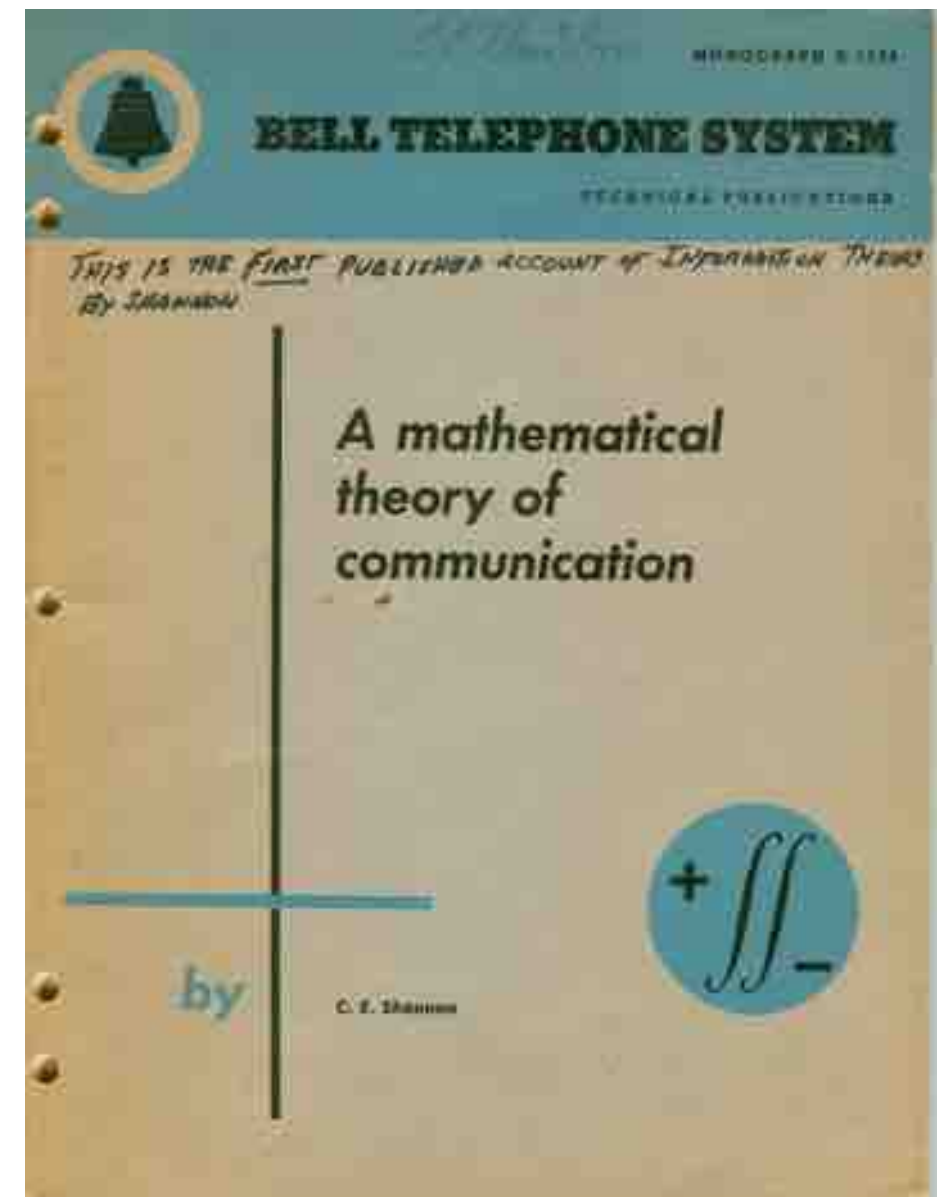
It showed how to make communications faster and take up less space on a hard disk, making the internet possible


$$H = -\sum p(x) \log p(x)$$

Введена мера информации(!)

кг, метр, секунда + БИТ

The Guardian

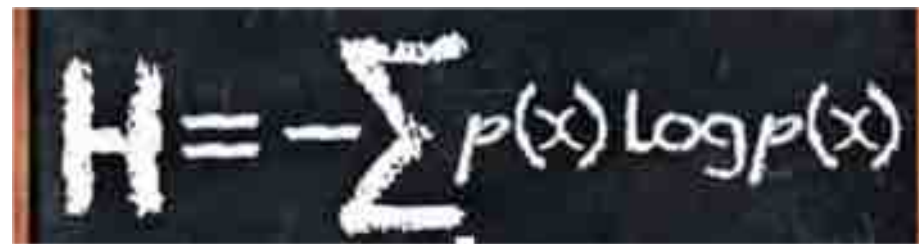


1948

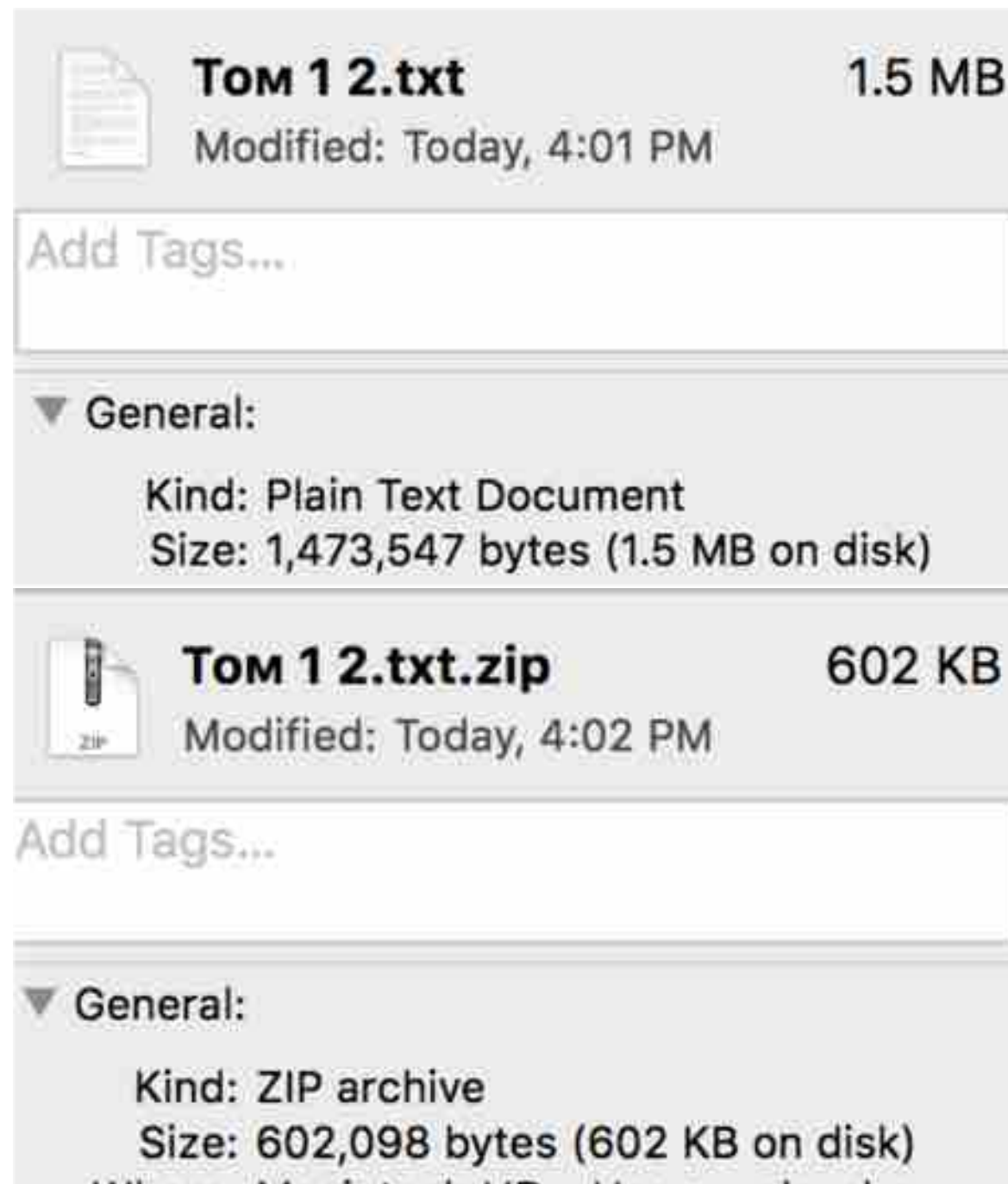
Информационная энтропия

[Клод Шеннон](#) предположил, что прирост информации равен утраченной неопределённости, и задал требования к её измерению:

1. мера должна быть непрерывной; то есть изменение значения величины вероятности на малую величину должно вызывать малое результирующее изменение функции;
2. в случае, когда все варианты (буквы в приведённом примере) равновероятны, увеличение количества вариантов (букв) должно всегда увеличивать значение функции;
3. должна быть возможность сделать выбор (в нашем примере букв) в два шага, в которых значение функции конечного результата должно являться суммой функций промежуточных результатов.


$$H = -\sum p(x) \log p(x)$$

Сжатие информации



Буква в тексте не независимы – одни встречаются чаще других в разных контекстах.

Сжатие информации

Кодирование длин серий

WWWWWWWWWWWWBWWWWWWWWWWWWBWWWWWWWWWWWWWWWWWWWWWWWWBWWWWWWWWWWWWWWWW

Посчитаем количество повторяющихся символов:

1. 12 символов «W»;
2. 1 символ «B»;
3. 12 символов «W»;
4. 3 символа «B»;
5. 24 символа «W»;
6. 1 символ «B»;
7. 14 символов «W».

Итого найдено 7 серий. Заменяем серии на число повторов и сам повторяющийся символ:

12W1B12W3B24W1B14W

Информационная энтропия

$$H = -\sum p(x) \log p(x)$$



$$H = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{1}{3} \log_2 \frac{1}{3} - \frac{1}{3} \log_2 \frac{1}{3} = -\log_2 \frac{1}{3}$$

1.5849625007211563... бит

Энтропия

Информационная и физическая энтропии
имеют глубинную связь



$$H \stackrel{\text{def}}{=} \int P(\ln P) d^3 v = \langle \ln P \rangle$$

В [термодинамике](#) и [кинетической теории](#), H - теорема, полученная [Больцманом](#) в [1872 году](#), описывает неубывания энтропии идеального газа в необратимых процессах, исходя из [уравнения Больцмана](#).

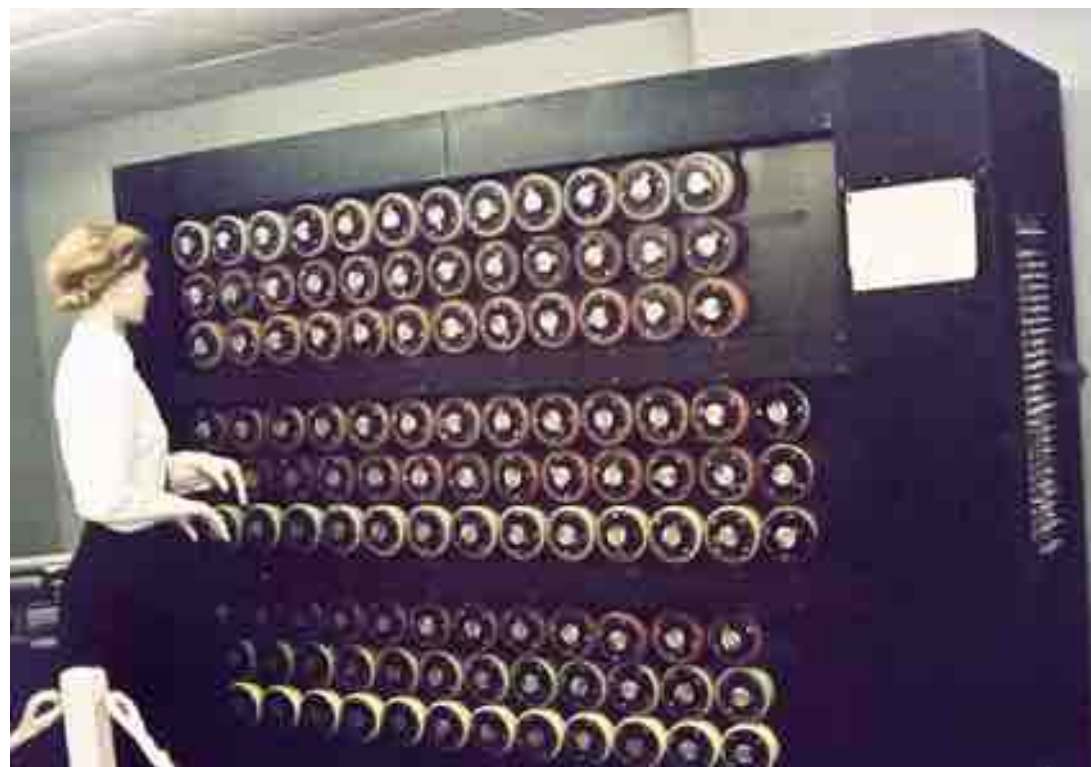
$$S \stackrel{\text{def}}{=} - NkH,$$

Ludwig Eduard Boltzmann

Передача информации

Передача информации

- Комплексная область: **теоретические, практические**, физические аспекты
- Вопросы сжатия данных
- Вопросы надежности
- Вопросы **шифрования и защиты данных** (особенно в медицине и биологии)



Передача информации

Связь частоты сигнала и пропускной способности

Владимир Александрович
Котельников



Harry Nyquist

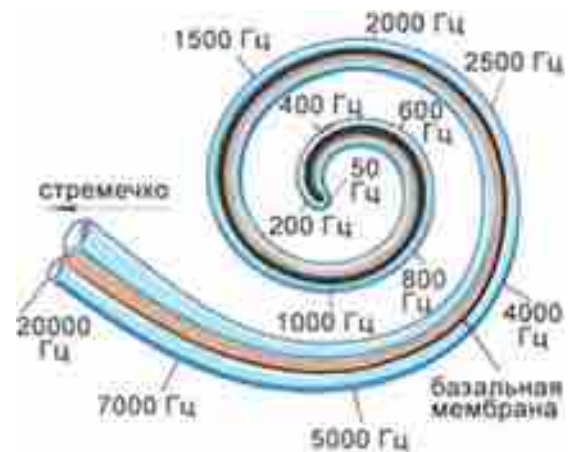
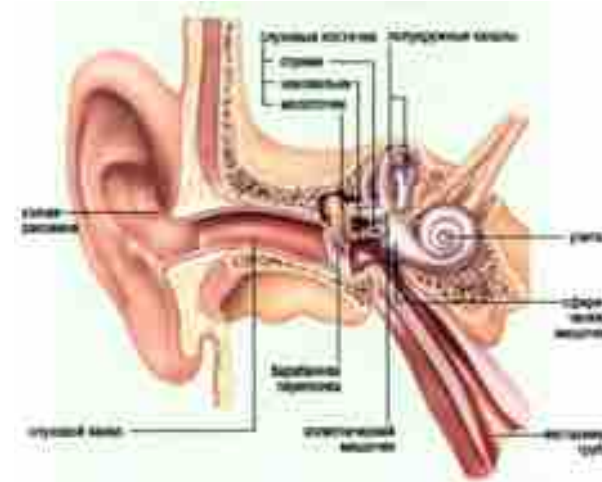


1908 - 2005

1889 – 1976

Теорема Котельникова-(Найквиста-Шенона)

«любую функцию $F(t)$, состоящую из частот от 0 до f , можно непрерывно передавать с любой точностью при помощи чисел, следующих друг за другом через $1/(2f)$ секунд



44.1 кГц – частота дискретизации при записи звука

Передача информации

Связь частоты сигнала и пропускной способности

Ralph Hartley



1888 – 1970

$$C = B \log_2 \left(1 + \frac{S}{N} \right),$$

где

C — пропускная способность канала, бит/с;

B — полоса пропускания канала, Гц;

S — полная мощность сигнала над полосой пропускания, Вт или B^2 ;

N — полная шумовая мощность над полосой пропускания, Вт или B^2 ;

S/N — отношение мощности сигнала к шуму (SNR).

Теорема Шеннона-Хартли

Передача информации

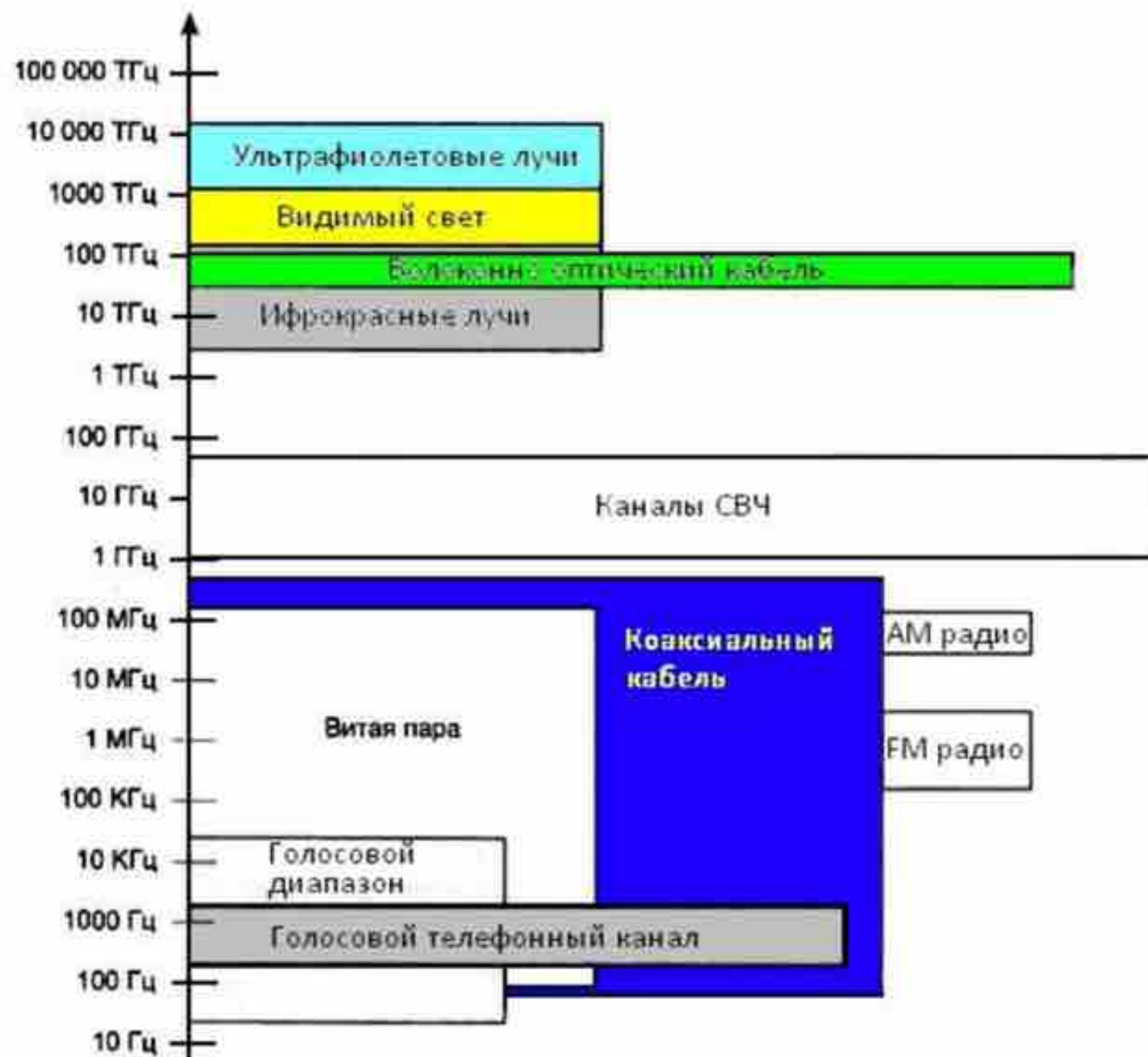
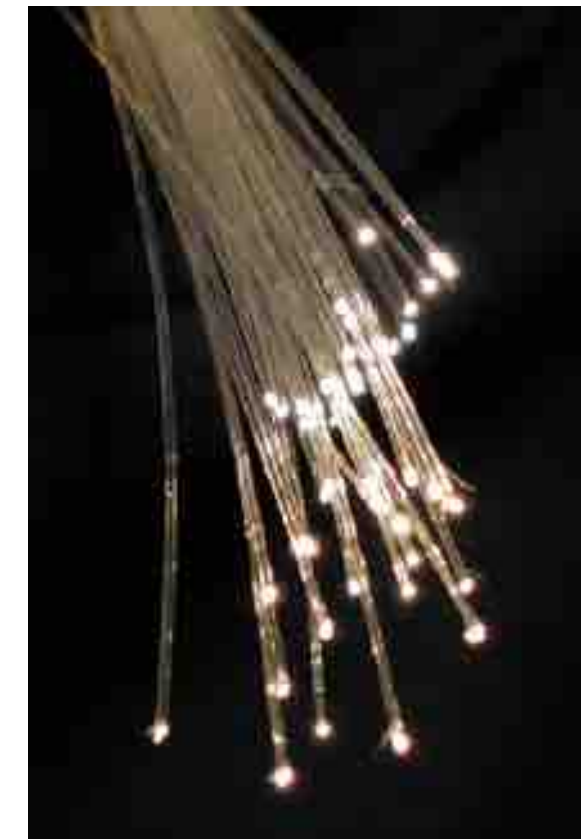


Рис. 1. Полосы пропускания линий связи и популярные частотные диапазоны

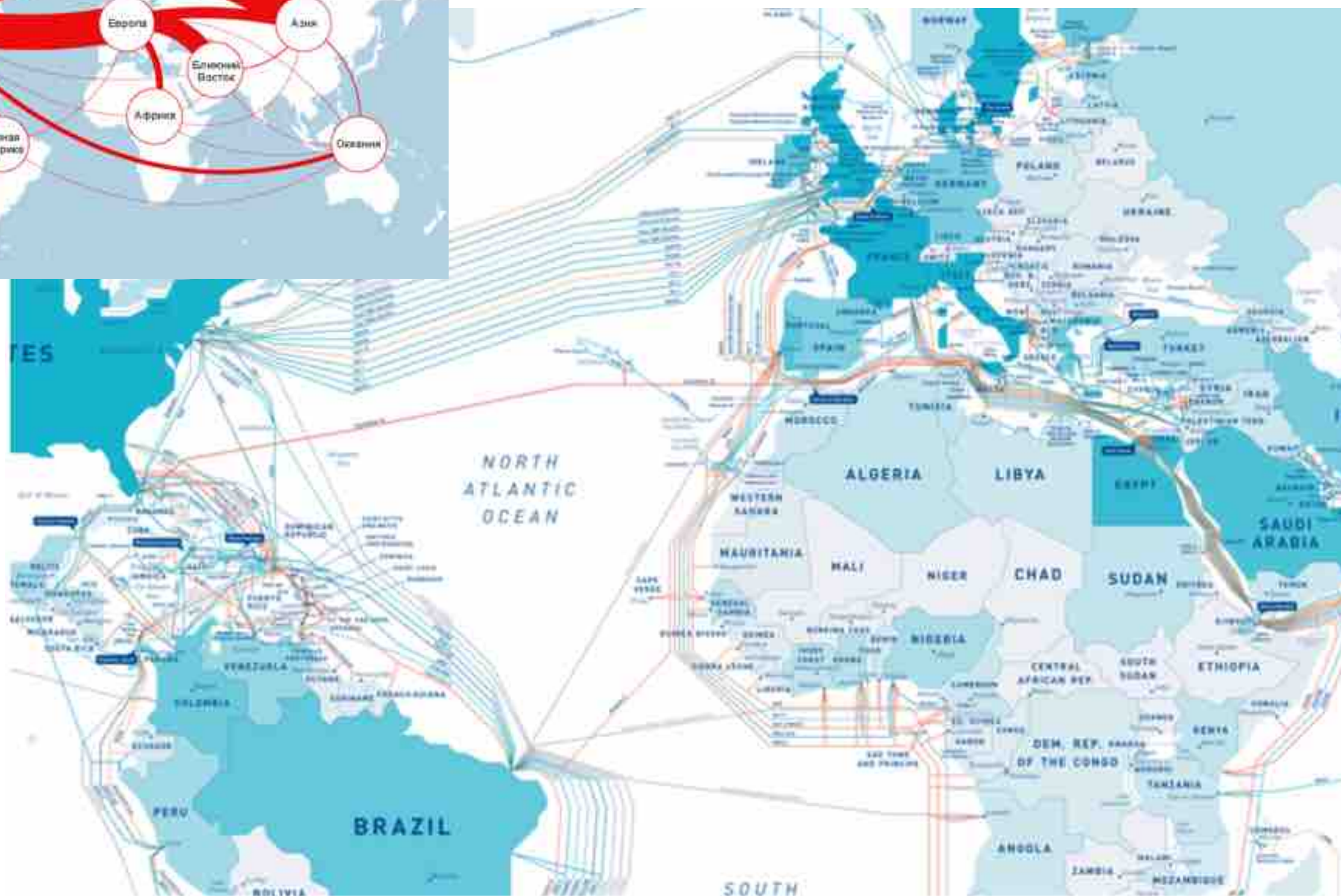
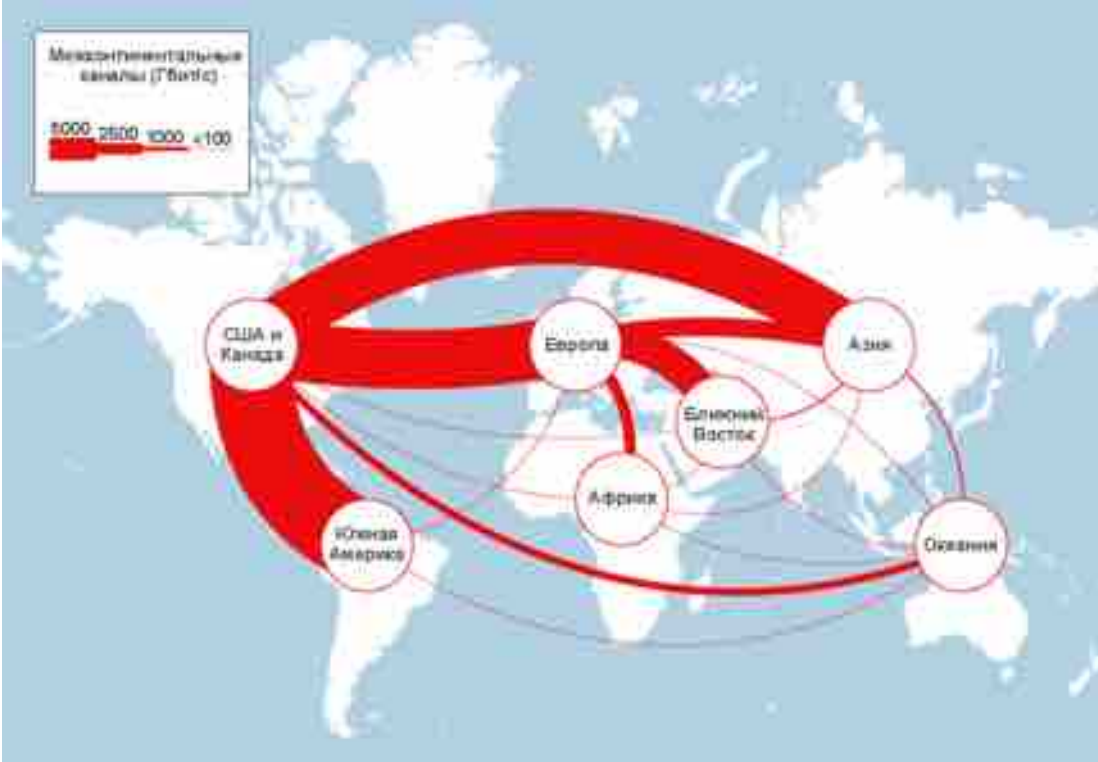


Оптоволокно



Антенны КНЧ

Каналы связи

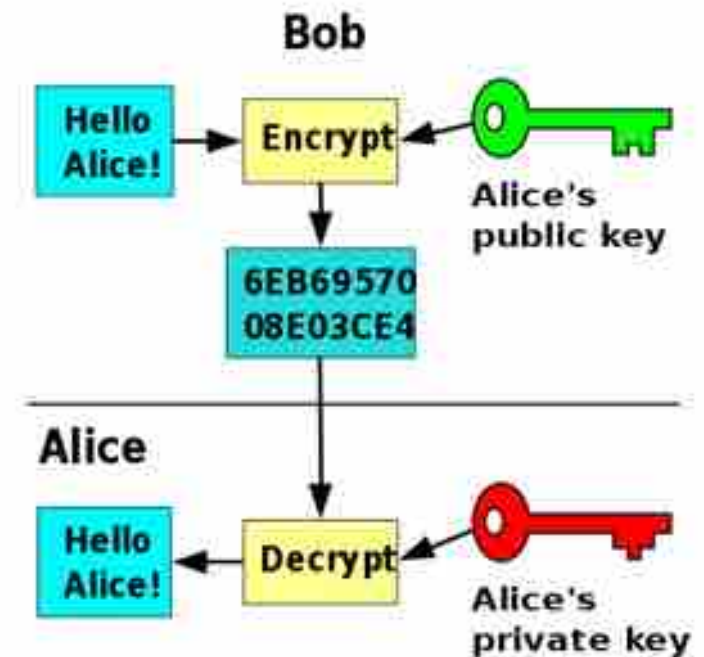


Карта подводных кабелей

Шифрование информации



Криптосистемы с открытым ключом



Необратимая Хэш функция

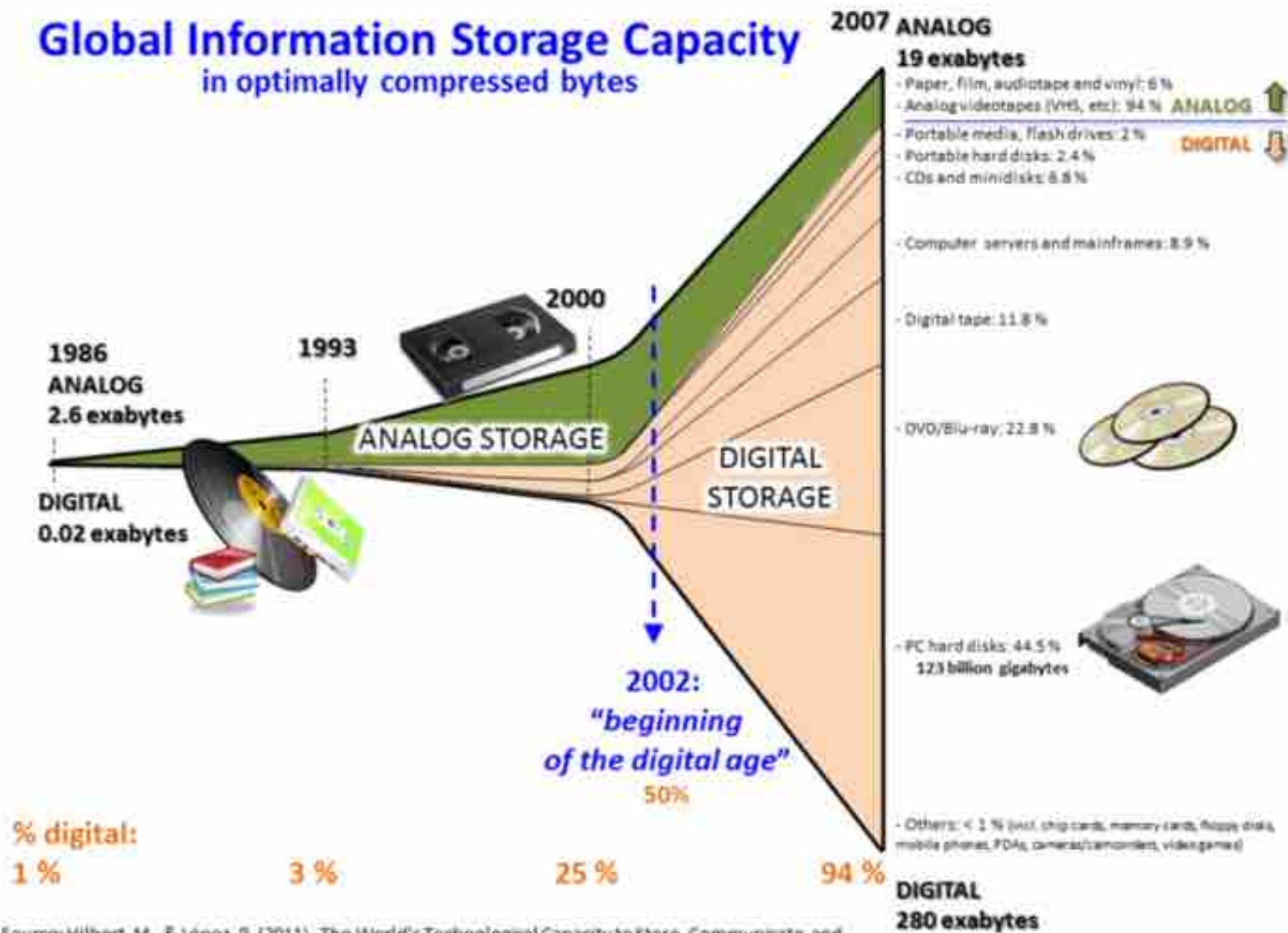
```
mbptb:~ alexsha$ md5 -s 'Hello world!!!'  
MD5 ("Hello world!!!") = 87ee732d831690f45b8606b1547bd09e
```

Хранение информации

Хранение информации

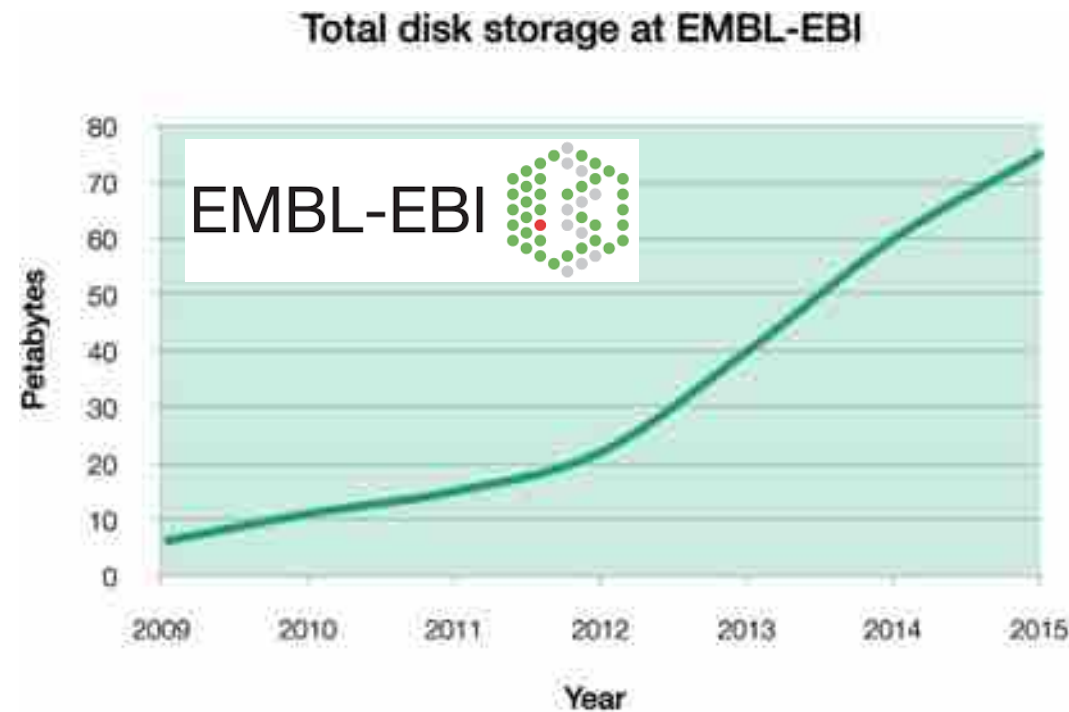


Хранение информации



Source: Hilbert, M., & López, P. (2011), The World's Technological Capacity to Store, Communicate, and Compute Information. Science, 332(6025), 60 –65. <http://www.martinhilbert.net/WorldInfoCapacity.html>

Большие данные в биомедицине



2013-2021
~\$400 млн

Table 1. Four domains of Big Data in 2025. In each of the four domains, the projected annual storage and computing needs are presented across the data lifecycle.

Data Phase	Astronomy	Twitter	YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

Источники больших данных в биомедицине

- **Омиксные технологии**

- Секвенирование, геномика, транскриптомика, протеомика, метаболомика и т.д.
- Коннектом мозга

- **Медицинская информация**

- Электронные медицинские карты, результаты клинических исследований и т.д.
- Медицинские изображения, МРТ и т.д.

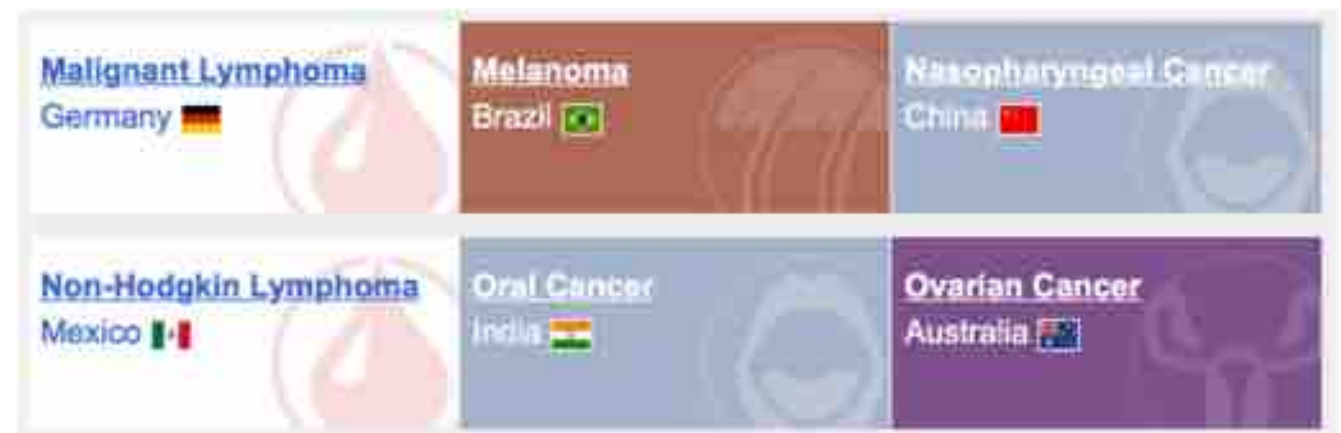
- **Структурная биология и моделирование**

- Данные с лазеров на свободных электронах (XFEL)
- Моделирование структуры и динамики белков.

Данные секвенирования, пример Геномы раковых опухолей



Геном человека ~ 3.3 Gb
x100 секвенирование ~300Gb

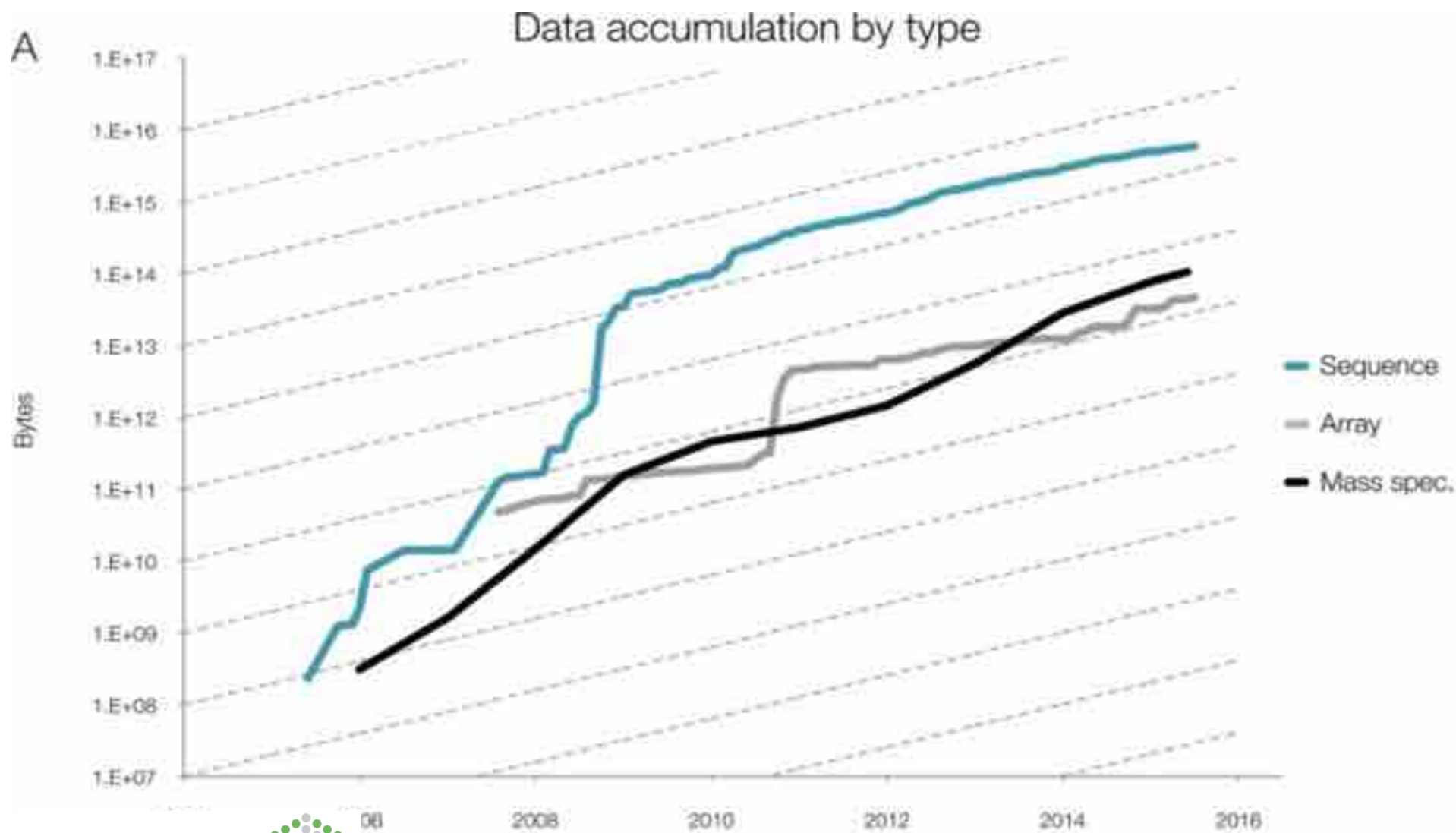
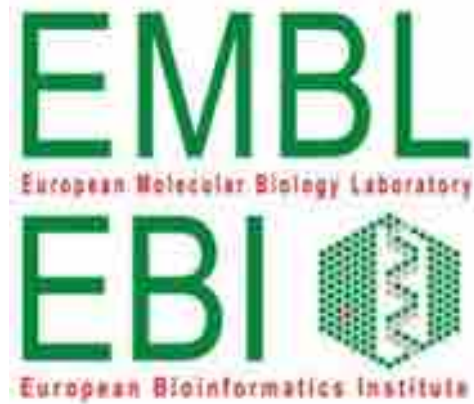


25000 образцов опухолей

Международный проект, данные распределены по миру

File ID	Donor	Repository	Project	Study	Data Type	Strategy	Format	Size	Link
F19955	DO217962	PCAWG - London, PCAWG - Barcelona, Collaboratory - Toronto, EGA - Hinxton	BRCA-EU	PCAWG	Aligned Reads	WGS	BAM	128.72 GB	Link
F19954	DO217962	PCAWG - London, PCAWG - Barcelona, Collaboratory - Toronto, EGA - Hinxton	BRCA-EU	PCAWG	Aligned Reads	WGS	BAM	107.27 GB	Link
F19974	DO46390	PCAWG - London, PCAWG - Barcelona, EGA - Hinxton, Collaboratory - Toronto, AWS - Virginia	OV-AU	PCAWG	Aligned Reads	WGS	BAM	134.05 GB	Link
F19973	DO46390	PCAWG - London, PCAWG - Barcelona, Collaboratory - Toronto, AWS - Virginia, EGA - Hinxton	OV-AU	PCAWG	Aligned Reads	WGS	BAM	101.45 GB	Link
F19956	DO222303	PCAWG - Chicago (TCGA), PDC - Chicago	DLBC-US	PCAWG	Aligned Reads	WGS	BAM	202.00 GB	
F19955	DO222303	PCAWG - Chicago (TCGA), PDC - Chicago	DLBC-US	PCAWG	Aligned Reads	WGS	BAM	100.77 GB	

Централизованные репозитории омиксных данных



European Bioinformatics Institute

Genomes en masse



5 years ~ 100 000
genomes



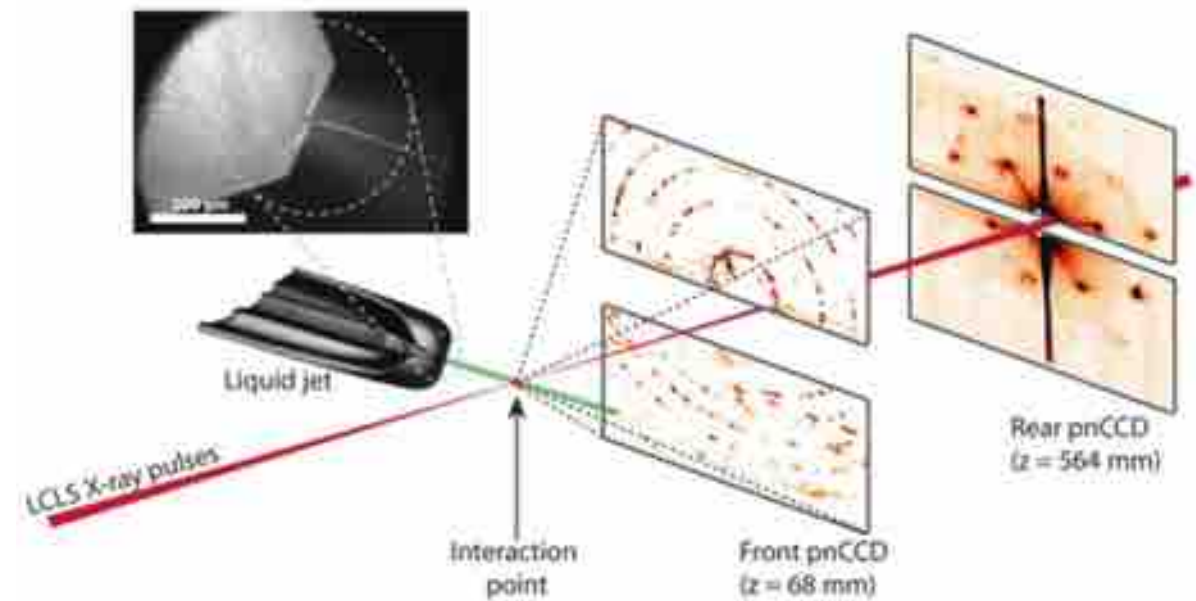
华大基因
BGI



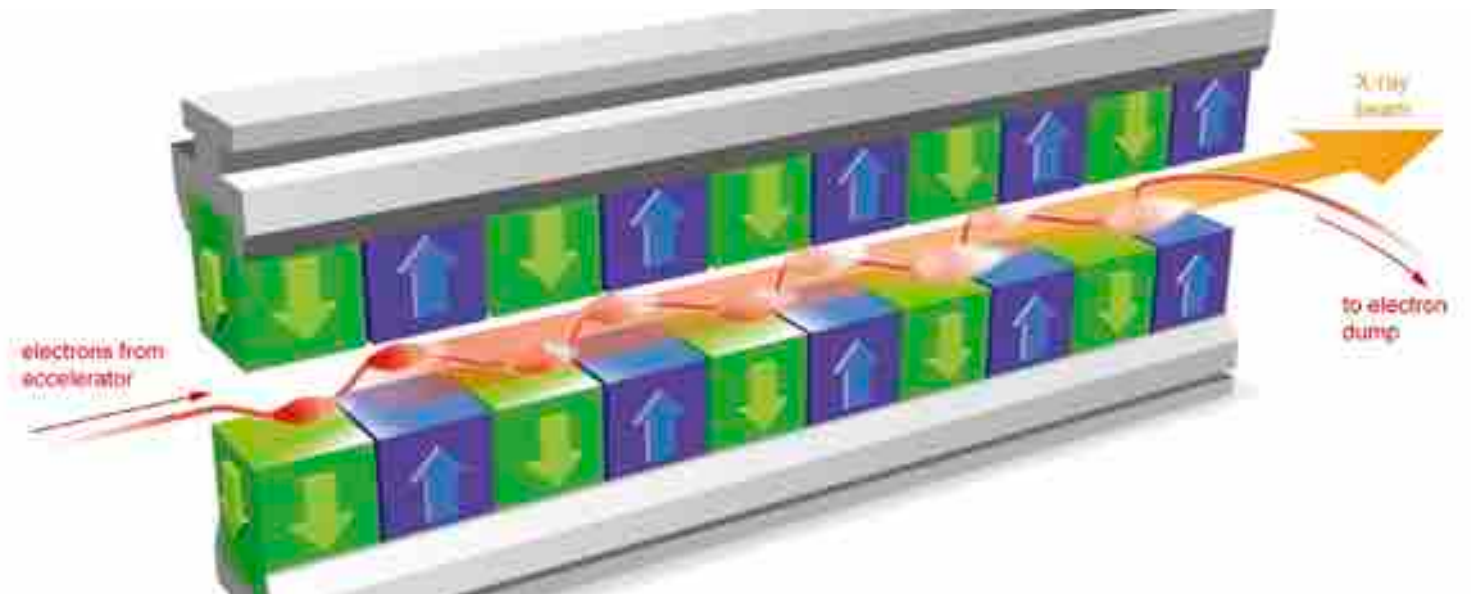
Структурная биология и моделирование



European XFEL, Hamburg



27000 импульсов в секунду



Передача больших данных

Выделенные научно-образовательные сети 100Gbit/s

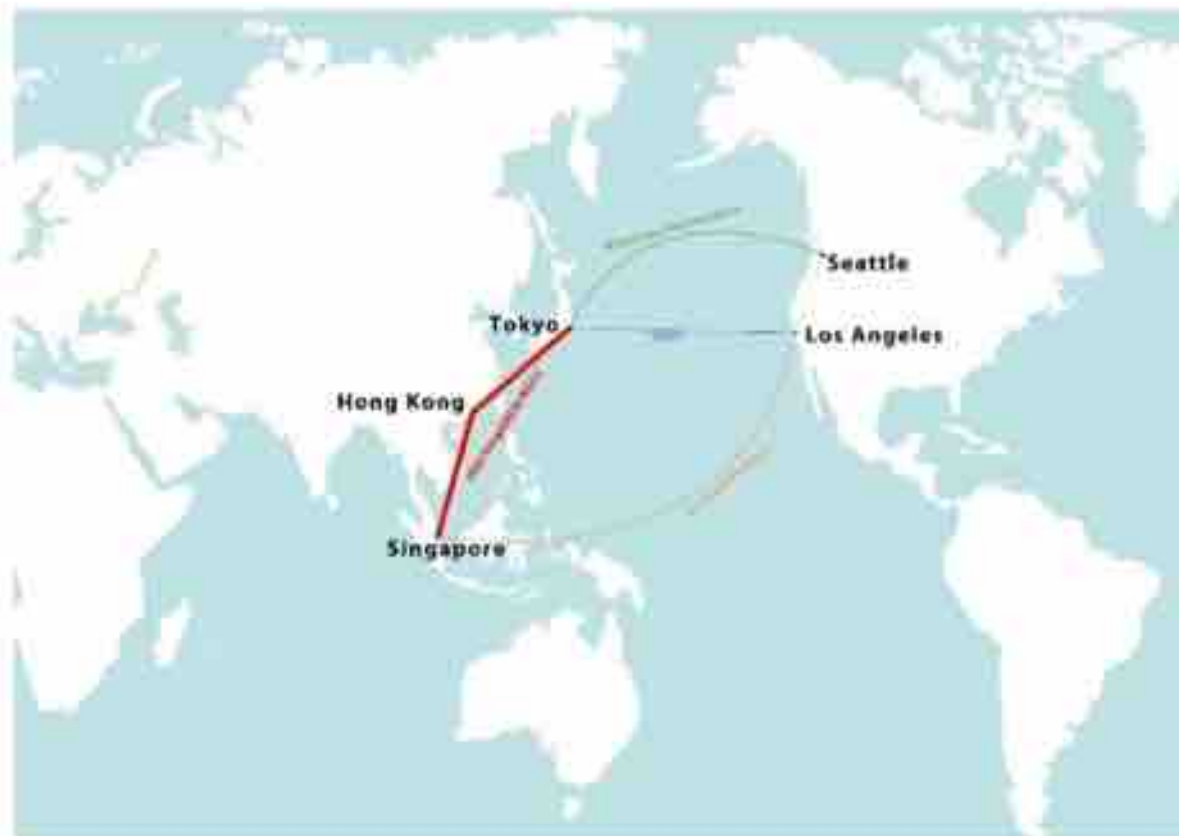


Figure 1: Asia Pacific Ring (APR)

INTERNET²



Программные решения



Базы данных

Базы данных

- Реляционные базы данных, объектно-ориентированные, RDF
- Системы управления базами данных СУБД
- Языки и стандарты SQL, SPARQL, RDF

Larry Ellison



Larry Ellison in 2016

Oracle Corporation

ORACLE®



Реляционные базы данных

Клиент

Id_кл	Фамилия	Имя	Отчество
15	Иванов	Иван	Иванович
16	Петров	Петр	Петрович
17	Николаев	Николай	Николаевич

Товар

Id_тов	Название
1	Шкаф
2	Стул
3	Стол

Заказ

Id_зак	Клиент	Товар	Дата	Количество
1	15	1	15.09.2003	1
2	17	1	17.09.2003	2
3	15	2	20.09.2003	12

SQL

Целостность
данных

Транзакции

Соответствие
требованиям ACID

Атомарность

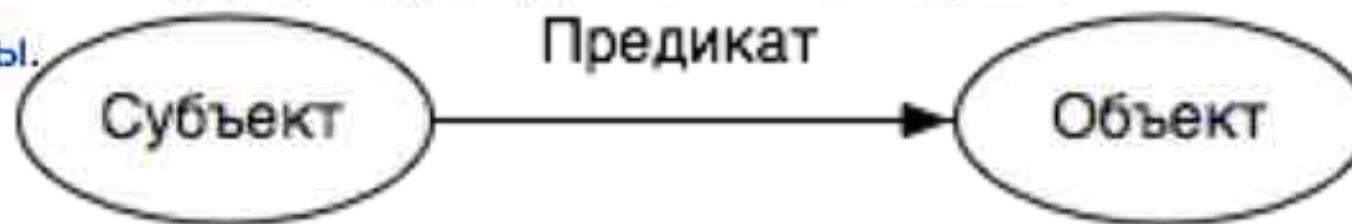
Изолированность

Надежность

Resource Description Framework

Язык SPARQL

Resource Description Framework (RDF, «среда описания ресурса»^[1]) — это разработанная консорциумом Всемирной паутины модель для представления данных, в особенности — метаданных^[2]. RDF представляет утверждения о ресурсах в виде, пригодном для машинной обработки. RDF является частью концепции семантической паутины.



```
1 PREFIX up:<http://purl.uniprot.org/core/>
2 PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
3 PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
4 SELECT ?protein ?text
5 WHERE
6 {
7     ?protein a up:Protein .
8     ?protein up:organism taxon:9606 .
9     ?protein up:annotation ?annotation .
10    ?annotation a up:Natural_Variant_Annotation .
11    ?annotation rdfs:comment ?text .
12    FILTER (CONTAINS(?text, 'loss of function'))
13 }
14
```

Select all human UniProt entries with a sequence variant that leads to a 'loss of function'

Биологические базы данных

Biology is a data-intensive science!

- Нужно уметь хранить данные
- Нужно уметь обрабатывать данные
- Нужно уметь обмениваться данными
- Данные должны быть максимально открыты и доступны научному сообществу.
- Data provenance (“происхождение данных”)

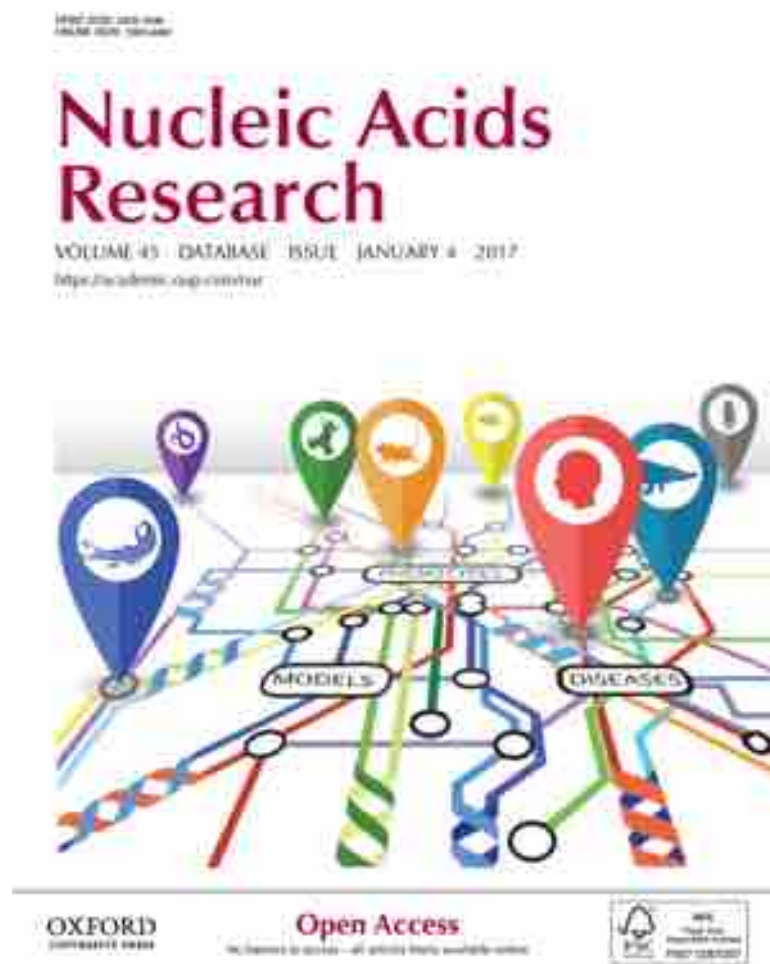
Data provenance [\[edit \]](#)

Scientific research is generally held to be of good provenance when it is documented in detail sufficient to allow reproducibility.^{[27][28]} Scientific workflow systems assist scientists and programmers with tracking their data through all transformations, analyses, and interpretations. Data sets are reliable when the process used to create them are

- Кризис воспроизводимости результатов в науке!?
-

Базы данных для биологии

- На данный момент количество не возможно сосчитать – очень много – важно не запутаться и не потеряться при их использовании
- Надежные источники информации о базах данных – научные журналы



Annual Database Issue – информация о ~200 БД каждый год.

Базы данных для биологии

- Бесплатные vs Платные (по подписке)
 - Свободно доступные vs Ограниченно доступные
 - Большие ресурсы (NCBI, EBI/EMBL, etc.)
интегрирующие многие базы данных -
поддерживаются государством
 - Коллаборации между университетами (напр. PDB)
 - Коммерческие компании
 - Локальные базы данных, поддерживаемые силами научных групп
 - База данных vs Web Server – граница размыта.
 - Хорошие БД - информационные ресурсы с возможностями сложного поиска и моделирования.
-

Крупные центры биологических БД



- Bethesda, MD USA
- Более 60 БД включая PubMed, GenBank, DBGap, SRA



- European Bioinformatics Institute, Cambridge, UK + Switzerland

Что храниться?

- БД статей, абстрактов, патентов
- Последовательности ДНК
- Последовательности белков
- 3D структуры молекул
- Геномы
- Данные экспрессии
- Сырые данные с секвенаторов
- Информация о химических соединениях и их активности
- Информация о болезнях, информация о пациентах
- Информация о видах живых организмов
- Информация о метаболических и сигнальных путях
- Информация о взаимодействии молекул
- Много производной информации: базы гомологичных последовательностей, аннотация отдельных классов белков и т.д.

План

- **Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)**
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

Реферативные базы данных

Clinical/Biomedical

PubMed – US National Library of Medicine database (Medline); refers to >25M articles from 5600 biomedical journals, 1940s to present, with some older items, in medicine, nursing, dentistry, veterinary medicine, allied health & pre-clinical sciences

- bibliographic database with author-provided abstracts, added indexing terms from **MeSH** (Medical Subject Headings) thesaurus, & links to other resources

www.pubmed.gov



FREE

Реферативные базы данных

Clinical/Biomedical

Embase – European based, includes all of Medline (database behind PubMed) and more; > 29M records, >8,500 journals, 1940s to present; includes coverage of more basic science journals & pre-clinical topics - especially useful for drug pipeline information, biotechnology, medical devices, conference coverage, toxicology, health policy/management, & alternative/complementary medicine
EMTREE thesaurus includes almost twice as many terms as PubMed

<https://www.elsevier.com/solutions/embase-biomedical-research>

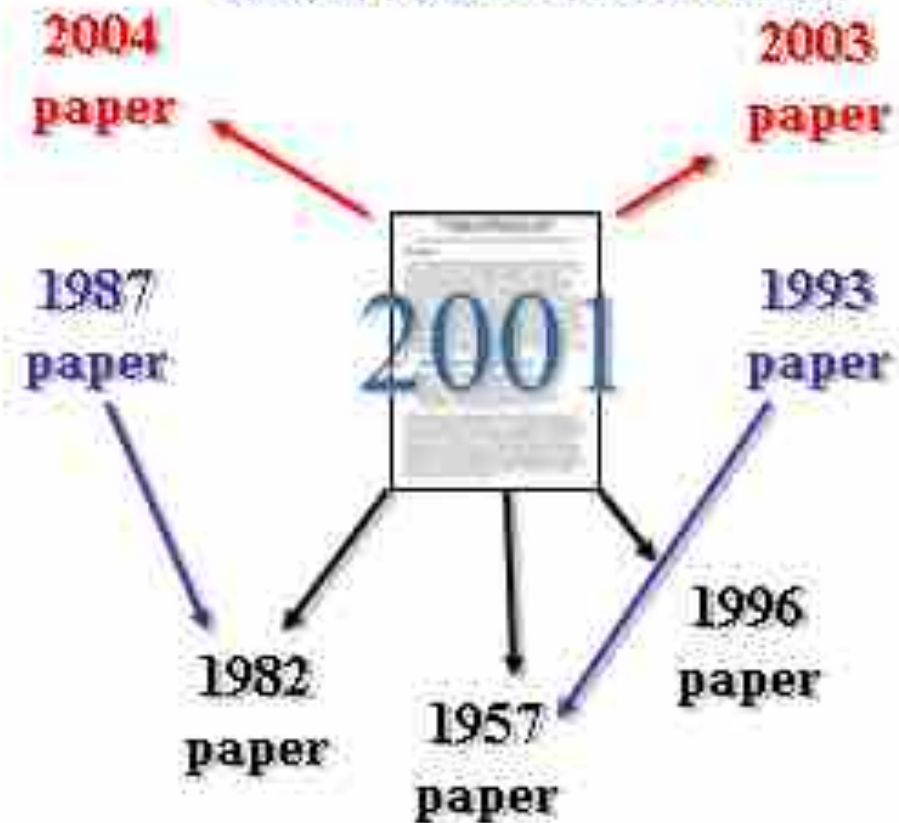
Реферативные базы данных

Cited Reference Searching

Traditional search



Cited reference search



Source: ISI

Реферативные базы данных

Общенаучные базы данных цитирований

Web of Science - covers >12,000 journals from 1900 to present;
useful for cited reference, **conference
information & affiliations** (institutions)

<https://webofknowledge.com/>



PAID

Scopus – covers >18,500 journals from 1823 to present, complete
citation counts for indexed articles 1996 to
present; a general science database, not a specialized database –
useful for cited reference, **conference
information & affiliations** (institutions)

<https://www.scopus.com/>

ELSEVIER



PAID

Реферативные базы данных

Общенаучные базы данных цитирований



НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА

eLIBRARY.RU

ЧИТАТЕЛЯМ | ОРГАНИЗАЦИЯМ | ИЗДАТЕЛЬСТВАМ | АВТОРАМ | БИБЛИОТЕКАМ

ОСНОВНЫЕ ПРОЕКТЫ НА ПЛАТФОРМЕ eLIBRARY.RU

Научная электронная библиотека eLIBRARY.RU - это крупнейший российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты более 26 млн научных статей и публикаций, в том числе электронные версии более 5600 российских научно-технических журналов, из которых более 4800 журналов в открытом доступе. [Подробнее...](#)

РОССИЙСКИЙ ИНДЕКС НАУЧНОГО ЦИТИРОВАНИЯ

Национальная библиографическая база данных научного цитирования, аккумулирующая более 12 миллионов публикаций российских ученых, а также информацию о цитировании этих публикаций из более 6000 российских журналов

SCIENCE INDEX ДЛЯ ОРГАНИЗАЦИЙ

Информационно-аналитическая система Science Index для анализа публикационной активности и цитируемости научных организаций

SCIENCE INDEX ДЛЯ АВТОРОВ

Инструменты и сервисы, предлагаемые для зарегистрированных авторов научных публикаций

RUSSIAN SCIENCE CITATION INDEX

Совместный проект компаний Clarivate Analytics и Научная электронная библиотека eLIBRARY.RU -

НОВОСТИ И ОБЪЯВЛЕНИЯ

- 26.01 Опубликована программа конференции SCIENCE ONLINE 2018
- 19.01 Открыта регистрация на семинар: "Использование РИНЦ и Science Index для анализа и оценки научной деятельности" 27 февраля
- 12.01 Открыт свободный доступ к архивам журналов Российской академии наук
- 05.10 Продолжение конкурсной поддержки программ развития научных журналов

[Другие новости](#)

ТЕКУЩЕЕ СОСТОЯНИЕ

Число наименований журналов:	62015
- из них российских журналов:	15351
- из них выходящих в настоящее время:	13115
Число журналов, индексируемых в РИНЦ:	5200
Число журналов с полными текстами:	10721
- из них в открытом доступе:	5537



Российская Академия Наук

Открытый доступ к журналам Российской академии наук



Подписка научных организаций на информационно-аналитическую систему

SCIENCE INDEX

поиск

Найти

[Расширенный поиск](#)

вход

IP-адрес компьютера:
128.68.143.128

Название организации:
[не определена](#)

Имя пользователя:
ashaytan

Пароль:

Elibrary.ru/РИНЦ

Реферативные базы данных

Общенаучные базы данных цитирований

Google Scholar



☒ Articles ☐ Case law

Recommended articles

Exploring DNA dynamics within oligonucleosomes with coarse-grained simulations:
SIRAH force field extension for protein-DNA complexes

A Brandner, A Schüller, F Melo, S Pantano - Biochemical and biophysical research ..., 2017

Базы данных диссертаций

Open DOAR <http://www.opendoar.org/index.html> ;

OpenThesis <http://www.openthesis.org/> ;

BASE – Bielefeld Academic Search Engine -
<http://www.base-search.net/>

> refine search result > document type > theses

ProQuest Dissertations & Theses

Database <http://www.proquest.com/products-services/pqdt.html> - from 1743

to present; some fulltext since 1990; **fee** with some
free search capability

Базы данных патентов



http://www.lens.org/lens/biological_search – поиск
ДНК последовательностей



План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- **Базы данных последовательностей ДНК**
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

Базы данных нуклеотидных последовательностей

Нуклеотидные БД – это хранилища, принимающие данные от научного сообщества и представляющие их широкой общественности. Различные БД отличаются по источнику последовательностей, их надежности, широте аннотирования и т.д. В идеале БД должна содержать все известные последовательности.

The **International Nucleotide Sequence Database Collaboration** – совместный проект EMBL-Bank в Европейском Институте Биоинформатики (EBI), японского банка данных ДНК (DDBJ) в Центре Информационной Биологии (CIB) и GenBank в Национальном Центре Биотехнологической Информации (NCBI).



База данных GenBank

Открытая БД нуклеотидных последовательностей, учреждена в 1982 г.

2017: > 300 000 организмов, ~ 203 млн. последовательностей,

~ 240 млрд. пар оснований



Sample GenBank Record

This page presents an annotated sample GenBank record (accession number **U49845**) in its *GenBank Flat File* format. You can see the corresponding [live record for U49845](#), and see [examples of other records](#) that show a range of biological features.

<u>LOCUS</u>	<u>SCU49845</u>	<u>5028 bp</u>	<u>DNA</u>	<u>PLN</u>	<u>21-JUN-1999</u>
<u>DEFINITION</u>	Saccharomyces cerevisiae TCP1-beta gene, partial cds, and Axl2p (AXL2) and Rev7p (REV7) genes, complete cds.				
<u>ACCESSION</u>	U49845				
<u>VERSION</u>	U49845.1 GI:1293613				
<u>KEYWORDS</u>	.				
<u>SOURCE</u>	Saccharomyces cerevisiae (baker's yeast)				
<u>ORGANISM</u>	Saccharomyces cerevisiae Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
<u>REFERENCE</u>	1 (bases 1 to 5028)				
<u>AUTHORS</u>	Torpey, L.E., Gibbs, P.E., Nelson, J. and Lawrence, C.W.				
<u>TITLE</u>	Cloning and sequence of REV7, a gene whose function is required for DNA damage-induced mutagenesis in Saccharomyces cerevisiae				
<u>JOURNAL</u>	Yeast 10 (11), 1503-1509 (1994)				
<u>PUBMED</u>	7871890				
<u>REFERENCE</u>	2 (bases 1 to 5028)				
<u>AUTHORS</u>	Roemer, T., Madden, K., Chang, J. and Snyder, M.				
<u>TITLE</u>	Selection of axial growth sites in yeast requires Axl2p, a novel plasma membrane glycoprotein				
<u>JOURNAL</u>	Genes Dev. 10 (7), 777-793 (1996)				
<u>PUBMED</u>	8846915				

База данных GenBank. Структура файла

...

```
FT      /translation="MGQPGNGSAFLLAPNGSHAPDHDVTQQRDEVWVVGMGIVMSLIVL
FT      AIVFGNVLVITAIKFERLQTVTNYFITSLACADLVMGLAVVPFGAAHILMKMWTFGNF
FT      WCEFWTSIDVLCVTASIETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSGL
FT      TSFLPIQMHWYRATHQEAINCYANETCCDFFTNQAYAIASSIVSFYVPLVIMVFVYSRV
FT      FQEAKRQLQKIDKSEGRFHVQNLSQVEQDGRGTGHGLRRSSKFCLKEHKALKTLGIIMGT
FT      FTLCWLPPFFIVNIVHVIQDNLIRKEVYILLNWIGYVNSGFNPLIYCRSPDFRIAFQELL
FT      CLRRSSLKAYGNGYSSNGNTGEQSGYHVEQEKENKLLCEDLPGTEDFVGHQGTVPSDNI
FT      DSQGRNCSTNDSL«
```

```
FT variation      46
```

```
FT      /gene="ADRB2«
```

```
FT      /replace="a«
```

```
FT      /note="Arg16 to Gly polymorphism«
```

```
XX
```

...

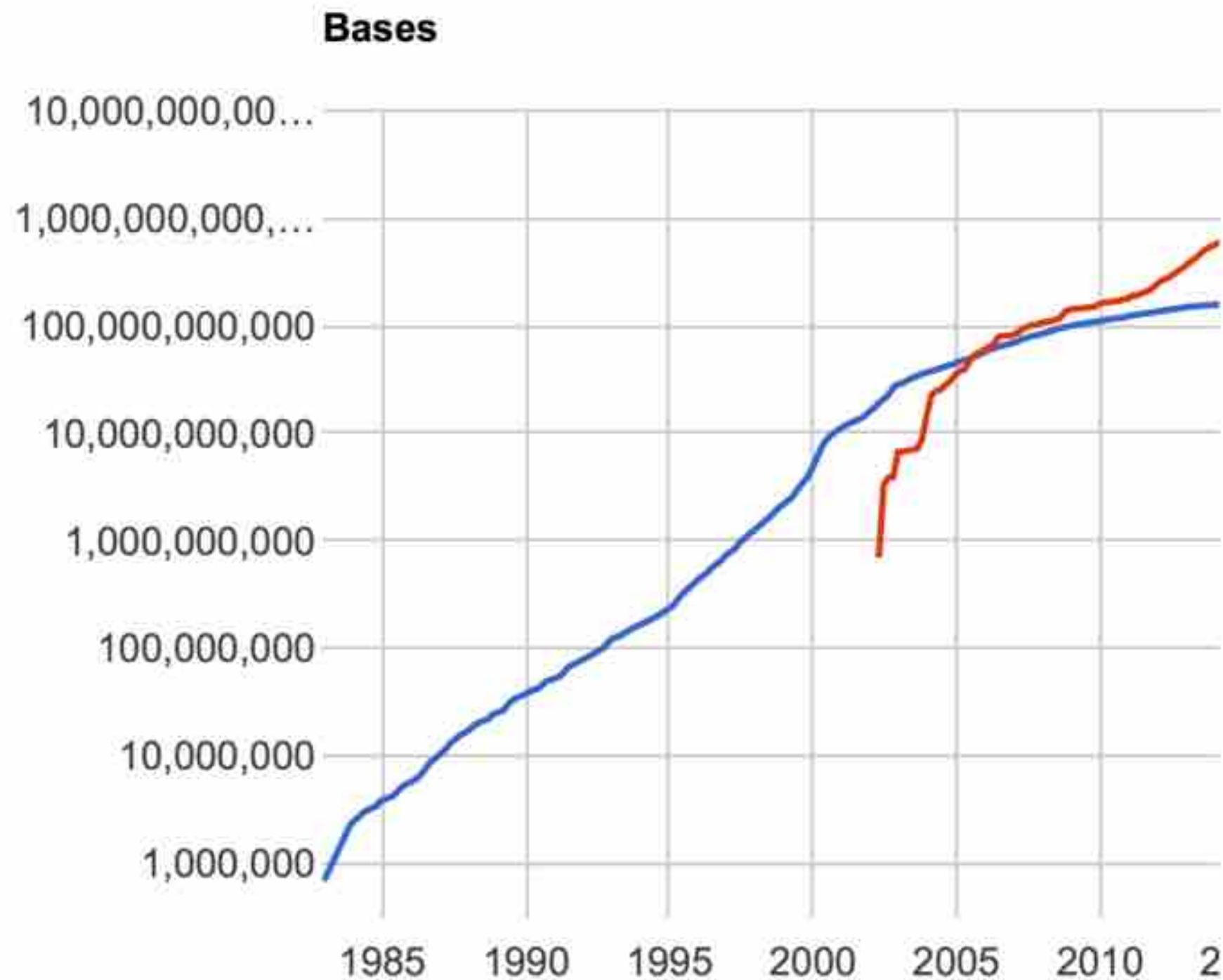
База данных GenBank. Структура файла

...

SQ Sequence 1242 BP; 275 A; 331 C; 326 G; 310 T; 0 other;

atggggcaac	ccgggaacgg	cagcgccttc	ttgctggcac	ccaatggaag	ccatgcgccc	60
gaccacgacg	tcacgcagca	aagggacgag	gtgtgggtgg	tgggcatggg	catcgtcatg	120
tctctcatcg	tcctggccat	cgtgtttggc	aatgtgctgg	tcatcacagc	cattgccaag	180
ttcgagcgtc	tgcagacggt	caccaactac	ttcatcactt	cactggcctg	tgctgatctg	240
gtcatggggc	tggcagtggg	gccctttggg	gccgcccata	ttcttatgaa	aatgtggact	300
tttggcaact	tctgggtgcga	gtttttggact	tccattgatg	tgctgtgcgt	cacggccagc	360
attgagaccc	tgtgcgtgat	cgcagtggat	cgctactttg	ccattacttc	acctttcaag	420
taccagagcc	tgctgaccaa	gaataaggcc	cgggtgatca	ttctgatggg	gtggattgtg	480
tcaggcctta	cctccttctt	gcccatcag	atgcactggg	accggggccac	ccaccaggaa	540
gccatcaact	gctatgccaa	tgagacctgc	tgtgacttct	tcacgaacca	agcctatgcc	600
attgcctctt	ccatcgtgtc	cttctacggt	cccctgggtga	tcattgggtctt	cgtctactcc	660
aggggtctttc	aggaggccaa	aaggcagctc	cagaagattg	acaaatctga	gggccgcttc	720
catgtccaga	accttagcca	ggtaggagcag	gatggggcgga	cgggggcatgg	actccgcaga	780
tcttccaagt	tctgcttgaa	ggagcacaaa	gccctcaaga	cgttagggcat	catcatgggc	840
actttcaccc	tctgctgggt	gcccttcttc	atcgттаaca	ttgtgcatgt	gatccaggat	900
aacctcatcc	gtaagggaagt	ttacatcctc	ctaaattgga	taggctatgt	caattctggg	960
ttcaatcccc	ttatctactg	ccggagccca	gatttcagga	ttgccttcca	ggagcttctg	1020
tgcttgcgca	gggtcttcttt	gaaggcctat	gggaatgggt	actccagcaa	cggcaacaca	1080
ggggagcaga	gtggatatca	cgtggaacag	gagaaagaaa	ataaactgct	gtgtgaagac	1140
ctcccaggca	cggaagactt	tgtgggccat	caagggtactg	tgcttagcga	taacattgat	1200
tcacaaggga	ggaattgtag	tacaaatgac	tcactgctgt	aa		1242

//



GenBank and WGS Statistics

<https://www.ncbi.nlm.nih.gov/genbank/statistics/>

Genbank – is an archive! Contains everything.

Nicotiana tabacum chloroplast JLA region, sequence 2

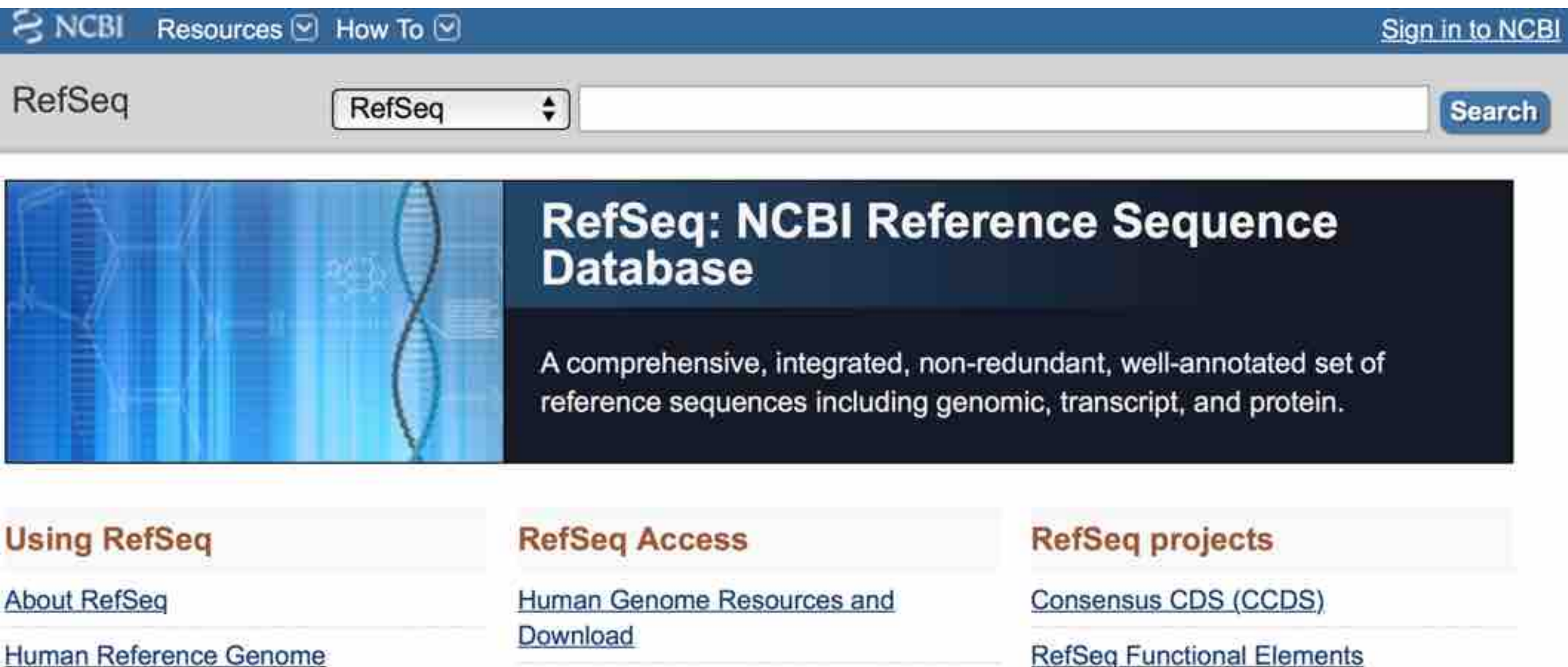
GenBank: Z71230.1

[FASTA](#) [Graphics](#)

FEATURES	Location/Qualifiers
source	1..124 /organism="Nicotiana tabacum" /organelle="plastid:chloroplast" /mol_type="genomic DNA" /isolate="Cuban cahibo cigar, gift from President Fidel Castro" /db_xref="taxon: 4097 "

RefSeq – is a reference sequence database!

RefSeq – is a reference sequence database!



The screenshot shows the top of the NCBI RefSeq website. At the top is a blue navigation bar with the NCBI logo, links for 'Resources' and 'How To', and a 'Sign in to NCBI' link. Below this is a search bar with 'RefSeq' entered in the dropdown menu and a 'Search' button. The main content area features a large banner with a blue background and a DNA double helix. The banner text reads 'RefSeq: NCBI Reference Sequence Database' and describes it as 'A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.' Below the banner are three columns of links. The first column, 'Using RefSeq', includes 'About RefSeq' and 'Human Reference Genome'. The second column, 'RefSeq Access', includes 'Human Genome Resources and Download'. The third column, 'RefSeq projects', includes 'Consensus CDS (CCDS)' and 'RefSeq Functional Elements'.

NCBI Resources How To Sign in to NCBI

RefSeq RefSeq Search

RefSeq: NCBI Reference Sequence Database

A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.

Using RefSeq

[About RefSeq](#)

[Human Reference Genome](#)

RefSeq Access

[Human Genome Resources and Download](#)

RefSeq projects

[Consensus CDS \(CCDS\)](#)

[RefSeq Functional Elements](#)

Если нужен список последовательностей всех генов человека – это вопрос к RefSeq, а не GenBank!

План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- **Базы данных последовательностей белков**
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

GenBank/RefSeq is nucleotide centric, but

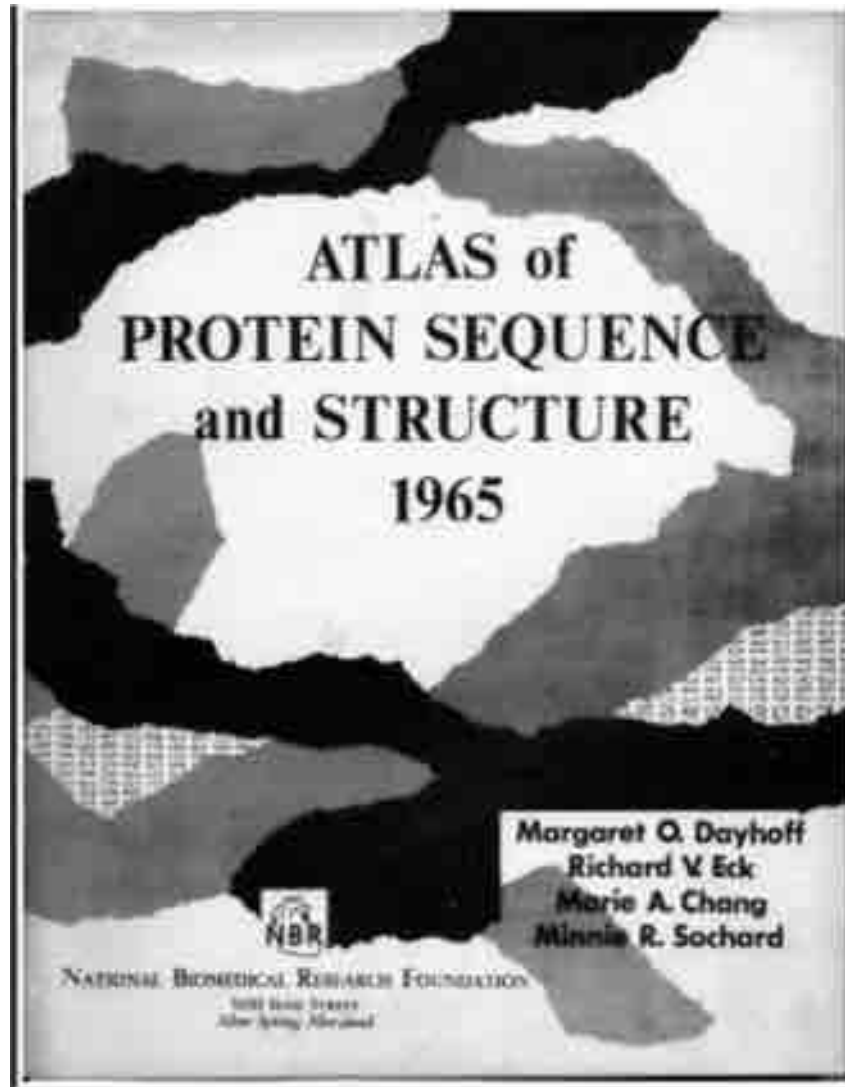
...

...

```
FT      /translation="MGQPGNGSAFLLAPNGSHAPDHDVTQQRDEVWVVGMGIVMSLIVL
FT      AIVFGNVLVITAIKFERLQTVTNYFITSLACADLVMGLAVVPFGAAHILMKMWTFGNF
FT      WCEFWTSIDVLCVTASIETLCVIAVDRIYFAITSPFKYQSLLTKNKARVILMVWIVSGL
FT      TSFLPIQMHWYRATHQEAINCYANETCCDFFTNQAYAIASSIVSFYVPLVIMVFVYSRV
FT      FQEAKRQLQKIDKSEGRFHVQNLSQVEQDGRTGHGLRRSSKFCLKEHKALKTLGIIMGT
FT      FTLCWLPFFIVNIVHVIQDNLIRKEVYILLNWIGYVNSGFNPLIYCRSPDFRIAFQELL
FT      CLRRSSLKAYGNGYSSNGNTGEQSGYHVEQEKENKLLCEDLPGTEDFVGHQGTVPDNI
FT      DSQGRNCSTNDSLL«
FT variation 46
FT      /gene="ADRB2«
FT      /replace="a«
FT      /note="Arg16 to Gly polymorphism«
XX
...
```

Protein sequences are annotated within GB records

Protein Centric Sequence Databases



Margaret Oakley Dayhoff

1925-1983

Margaret Dayhoff, a founder of the field of bioinformatics

Invented one-letter amino acid code, substitution matrices, etc.

https://en.wikipedia.org/wiki/Margaret_Oakley_Dayhoff

Protein Centric Sequence Databases



<http://pir.georgetown.edu>

In 2002, PIR along with its international partners, EBI (European Bioinformatics Institute) and SIB (Swiss Institute of Bioinformatics), were awarded a grant from NIH to create UniProt, a single worldwide database of protein sequence and function, by unifying the PIR-PSD, Swiss-Prot, and TrEMBL databases. As of 2010, PIR offers a wide variety of resources mainly oriented to assist the propagation and standardization of protein annotation: PIRSF,^[8] iProClass, and iProLINK.

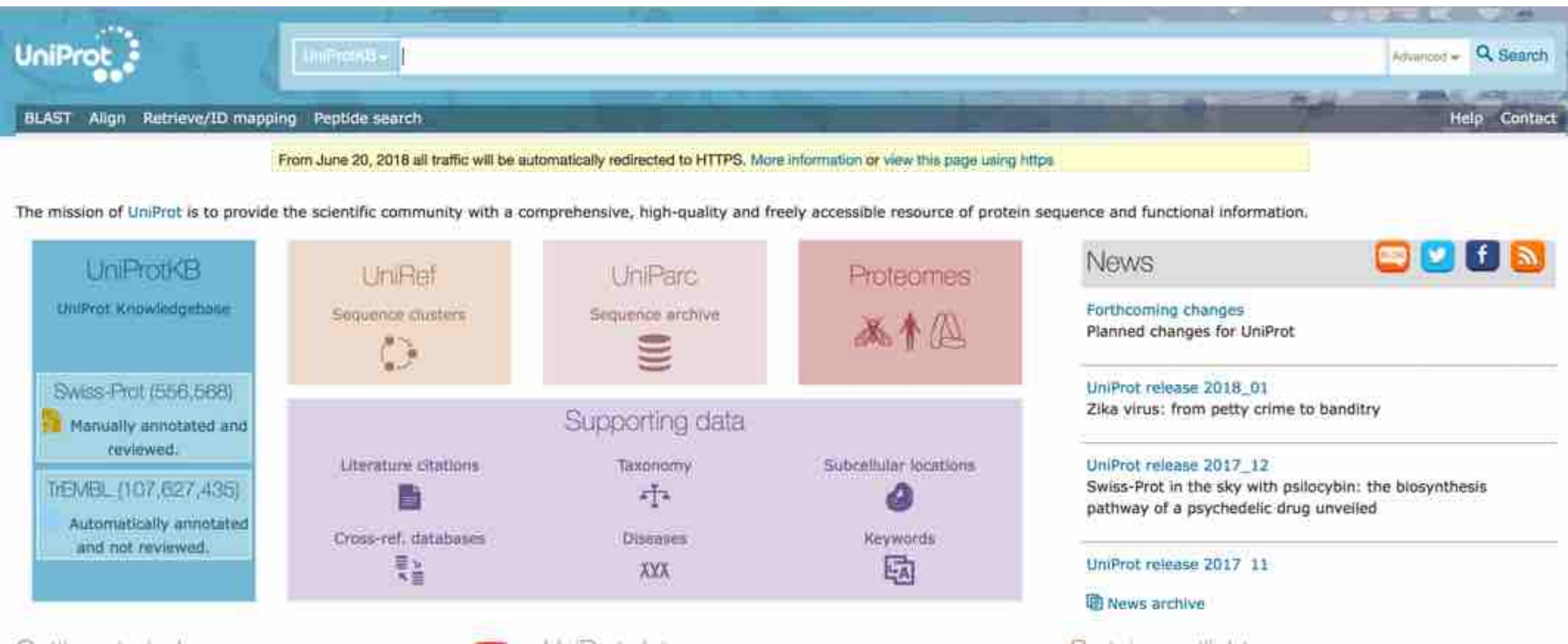
The Protein Ontology (PRO) is another popular database released by the Protein Information Resource.^{[9][10]}

Белковые базы данных



UniProt – наиболее всеобъемлющий каталог информации о белках, объединяющий в себе данные из UniProtKB/Swiss-Prot, UniProtKB/TrEMBL и PIR.

Белковые базы данных



Качественно аннотированную информацию о белках
нужно искать в **UniProtKB**

План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- **Базы данных 3D структур**
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

Структурные базы данных

The screenshot shows the RCSB PDB website interface. At the top, the logo for the Protein Data Bank (PDB) is displayed, along with the text "A MEMBER OF THE wwPDB". Below this, a banner states "An Information Portal to Biological Macromolecular Structures" and provides a timestamp: "As of Tuesday Oct 12, 2010 at 5 PM PDT there are 68562 Structures". A search bar is located in the top right, with a "Search" button and a link to "Advanced Search".

The main content area features a large heading "A Resource for Studying Biological Macromolecules". Below this heading, two paragraphs describe the PDB archive and the resources provided by the RCSB PDB. The first paragraph states: "The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the wwPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards." The second paragraph states: "The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists."

On the left side of the page, there are several navigation menus. The "MyPDB" menu includes links for "Login to your Account" and "Register a New Account". The "Home" menu includes links for "News & Publications", "Usage/Reference Policies", "Deposition Policies", "Website FAQ", "Deposition FAQ", "Contact Us", "About Us", "Careers", "External Links", "Sitemap", and "New Website Features". The "Deposition" menu includes links for "All Deposit Services", "Electron Microscopy", "X-ray | NMR", "Validation Server", "BioSync Beamline", and "Related Tools". A "Search" menu is also present.

On the right side of the page, there are several featured sections. The "Customize This Page" section includes a "New Features" menu with a link to "Improved Domain Display on Sequence Tab" and a "Website Release Archive" dropdown. The "RCSB PDB News" section includes a "Weekly | Quarterly | Yearly" filter and a "Molecule of the Month" section. The "Molecule of the Month" section features a "Categorized View of Molecule of the Month" and a "Molecule of the Month: Riboswitches" section. The "Featured Molecules" section includes a "Protein Synthesis" section and a "Molecule of the Month: Riboswitches" section.

PDB – содержит информацию об экспериментально определенных структурах белков, нуклеиновых кислот и различных комплексов.

Структурные базы данных

RCSB PDB
PROTEIN DATA BANK

На февраль 2016: 116 085 структур

1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016

2016-02-16 10:16:12
RCSB PDB
PROTEIN DATA BANK

Структурные базы данных

POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K

MyPDB

Hide

Login to your Account
Register a New Account

Home

Hide

News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

Deposition

Hide

All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Bearaliba
Related Tools

Search

Hide

Advanced Search
Latest Release
New Structure Papers
Sequence Search
Chemical Components
Unreleased Entries
Browse Database

Summary

Sequence

Derived Data

Seq. Similarity

3D Similarity

Literature

Biol. & Chem.

Methods

Geometry

Links

POTASSIUM CHANNEL KCSA IN COMPLEX WITH
TETRABUTYLAMMONIUM IN HIGH K

DOI:10.2210/pdb2jk5/pdb

2JK5

Display Files
Download Files
Print this Page
Share this Page

Primary Citation

Structure of a Potassium Channel in Complex with Symmetrical Quaternary Ammonium Compounds Reveals a Novel Hydrophobic Binding Site
Lenaeus, M.J., Focia, P.J., Wagner, T., Gross, A.
Journal: To be Published
Not in PubMed

Molecular Description

Hide

Classification: Immune System/metal Transport
Structure Weight: 61280.07

Molecule: ANTIBODY FAB FRAGMENT LIGHT CHAIN
Polymer: 1 Type: polypeptide(L) Length: 219
Chains: A

Molecule: ANTIBODY FAB FRAGMENT HEAVY CHAIN
Polymer: 2 Type: polypeptide(L) Length: 212
Chains: B

Molecule: VOLTAGE-GATED POTASSIUM CHANNEL
Polymer: 3 Type: polypeptide(L) Length: 124
Chains: C

Biological Assembly

Biological Assembly Image for 2JK5

More Images...

View in Jmol SimpleViewer
Other Viewers Protein Workshop

Biological assembly generated by PISA
(software)

Структурные базы данных

POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM IN HIGH K

Deposition

Hide

All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamline
Related Tools

Search

Hide

Advanced Search
Latest Release
New Structure Papers
Sequence Search
Chemical Components
Unreleased Entries
Browse Database
Histograms

Explorer:

Last Structure: 2JK5

Tools

Hide

File Downloads
Compare Structures
FTP Services
File Formats
Services: RESTful | SOAP
Widgets

Education

Hide

Understanding PDB Data
Molecule of the Month
Educational Resources

Chain Display

Chain A (polymer 1) [help] [fasta] [text/markup]

Description

ANTIBODY FAB FRAGMENT LIGHT CHAIN

Chain Type

polypeptide(L)

Length

219 residues

dssp secondary structure

5% helical (4 helices; 12 residues)
51% beta sheet (22 strands; 112 residues)

[hide] [reference]

More annotations

Select

Sequence Details

dssp

PDB

PDB

QVQLQQPGAELVKPGASVKLSCKASGYTFTSDWIHWVKQRPFGHGLEWIGEIIPSYGRANY

1102030405060

dssp

PDB

PDB

NEKIQKKATLTADKSSSTAFMQLSSLTSEDSAVYYCARERGDGYFAVWGAGTTVTVSSAK

61708090100110120

dssp

PDB

PDB

TTPPSVYPLAPGSAAQNSMVTLGCLVKGYFPEPVTVTWNSGSLSSGVHTFPAVLQSDLY

121130140150160170180

dssp

PDB

PDB

TLSSSVTVPSSSWPSETVTNVAHPASSTKVDKKIVPRD

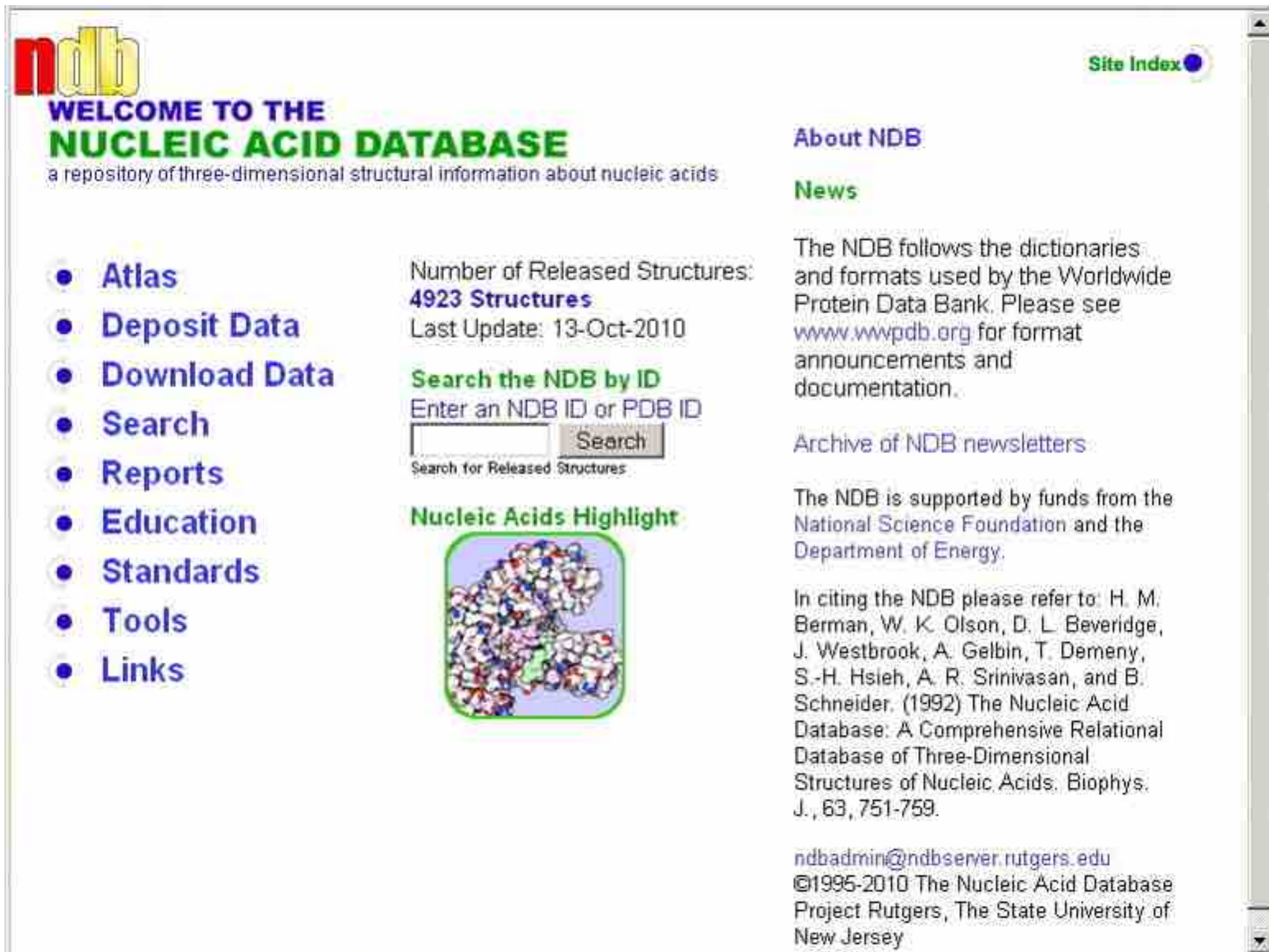
181190200210219

База данных PDB. Структура файла

```
HEADER      IMMUNE SYSTEM/METAL TRANSPORT          15-AUG-08      2JK5
TITLE       POTASSIUM CHANNEL KCSA IN COMPLEX WITH TETRABUTYLAMMONIUM
TITLE       2 IN HIGH K
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: ANTIBODY FAB FRAGMENT LIGHT CHAIN;
COMPND      3 CHAIN: A;
COMPND      4 ENGINEERED: YES;
COMPND      5 MOL_ID: 2;
...
KEYWDS      IMMUNE SYSTEM METAL TRANSPORT COMPLEX, QUATERNARY AMMONIUM,
...
EXPDTA      X-RAY DIFFRACTION
AUTHOR      M.J.LENAEUS,P.J.FOCIA,T.WAGNER,A.GROSS
REVDAT      1 17-NOV-09 2JK5 0
JRNL        AUTH      M.J.LENAEUS,P.J.FOCIA,T.WAGNER,A.GROSS
JRNL        TITL       STRUCTURE OF A POTASSIUM CHANNEL IN COMPLEX WITH
JRNL        TITL 2 SYMMETRICAL QUATERNARY AMMONIUM COMPOUNDS REVEALS
JRNL        TITL 3 A NOVEL HYDROPHOBIC BINDING SITE
JRNL        REF        TO BE PUBLISHED
JRNL        REFN
REMARK      2
REMARK      2 RESOLUTION.      2.4  ANGSTROMS.
REMARK      3
REMARK      3 REFINEMENT.
REMARK      3 PROGRAM      : REFMAC 5.5.0051
...
```

Структурные базы данных

NDB – основана в 1992 г. для сбора и распространения информации о структуре нуклеиновых кислот. Формат хранения данных идентичен PDB.



The screenshot shows the NDB website with a navigation menu on the left, a central search and statistics section, and informational text on the right. The navigation menu includes links to Atlas, Deposit Data, Download Data, Search, Reports, Education, Standards, Tools, and Links. The central section displays the number of released structures (4923) and the last update date (13-Oct-2010), along with a search box for NDB or PDB IDs. The right section provides information about the NDB, including its support by the National Science Foundation and the Department of Energy, and a list of authors for the database's creation.

WELCOME TO THE NUCLEIC ACID DATABASE
a repository of three-dimensional structural information about nucleic acids

- [Atlas](#)
- [Deposit Data](#)
- [Download Data](#)
- [Search](#)
- [Reports](#)
- [Education](#)
- [Standards](#)
- [Tools](#)
- [Links](#)

Number of Released Structures:
4923 Structures
Last Update: 13-Oct-2010

[Search the NDB by ID](#)
Enter an NDB ID or PDB ID

Search for Released Structures

Nucleic Acids Highlight



[Site Index](#)

[About NDB](#)

[News](#)

The NDB follows the dictionaries and formats used by the Worldwide Protein Data Bank. Please see www.wwpdb.org for format announcements and documentation.

[Archive of NDB newsletters](#)

The NDB is supported by funds from the National Science Foundation and the Department of Energy.

In citing the NDB please refer to: H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. (1992) The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids. *Biophys. J.*, 63, 751-759.

ndbadmin@ndbserver.rutgers.edu
©1995-2010 The Nucleic Acid Database
Project Rutgers, The State University of New Jersey

Структурные базы данных

EMBL-EBI

Find

Terms of Use | Privacy | Cookies

Databases | Tools | Research | Training | Industry | About Us | Help

Site Index

Home

EMD-1367

Mirrors

Contact us



Title: Three-dimensional structure of a voltage-gated potassium channel at 2.5 nm resolution.

Authors: Olga Sokolova, Ludmila Kolmakova-Partensky and Nikolaus Grigorieff

Sample: Shaker B channel

Aggregation state: Single particle (25 angstroms resolution)

Latest update: 2011-05-26

- Summary
- Experimental details
- Visualization
- Map information
- Downloads

Summary

Status: Released

Deposition date: 2007-05-24

Header release date: 2007-05-30

Map release date: 2007-05-30

Primary citation: Sokolova O, Kolmakova-Partensky L, Grigorieff N: Three-dimensional structure of a voltage-gated potassium channel at 2.5 nm resolution. *STRUCTURE* (2001) **9**, pp. 215-220 [PubMed 11286888]

Sample: Shaker B channel

Resolution: 25 Å (determined by FSC at 0.5 cut-off)

Fitted PDB:	PDB Authors	PubMed Status
1bl8	Doyle, D.A., Cabral, J.M., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., Mackinnon, R.	9525859 Released
1a68	Kreusch, A., Pfaffinger, P.J., Stevens, C.F., Choe, S.	9582078 Released

Структурные базы данных



Cambridge Crystallographic Data Centre

- About CCDC
- Products
- Case Studies
- Free Services
- News
- Events
- Contact
- Support
- Search

Location: Home

Products

Free Services

- Cambridge Structural Database (CSD)
- Deposit a Structure
- Request a Structure

New consultancy service

News

Last Updated:
10 September 2010

Positions Available

 **日本語**

500,000th Crystal Structure Milestone Reached

The Cambridge Crystallographic Data Centre (CCDC) is proud to announce the archiving of the 500,000th small molecule crystal structure to the Cambridge Structural Database (CSD)



"The determination of 500,000 crystal structures is a remarkable achievement. However, the scientific community is hungry for the next 500,000 and the knowledge these will undoubtedly bring. As the CSD grows both in size and in the complexity of structures it contains, the database not only helps us to answer our questions about molecular structure and interactions, but tells us what those questions should be."

-Dr Colin Groom, Executive Director of the CCDC

План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- **Базы данных хим. соединений**
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

Базы данных химических соединений

[Databases](#) | [Deposition](#) | [Services](#) | [Help](#) | [more](#)

PubChem

[BioAssay](#) [Compound](#) [Substance](#)

[GO](#) [Advanced search](#)

[Chemical structure search](#) | [BioActivity analysis](#)

New Structures from **Zancheng Functional Chemicals** are now available in PubChem.

[more ...](#) 

[Write to Helpdesk](#) | [Disclaimer](#) | [Privacy Statement](#) | [Accessibility](#)
National Center for Biotechnology Information
NLM | NIH | HHS

- [Bioactivity summary](#) 
- [Bioactivity datatable](#) 
- [Bioactivity structure-activity](#) 
- [Chemical structure search](#) 
- [3D conformer viewer](#) 
- [Chemical structure clustering](#) 
- [Deposition gateway](#) 
- [Structure download](#) 
- [PubChem FTP](#) 

Базы данных химических соединений

NCBI **PubChem** Compound

My NCBI [Sign In] [Register]

Search PubChem Compound for epinephrine Go Clear Save Search

Advanced Search Preview/Index History Clipboard Details

Display Summary Show 20 Sort By Send to

Tools: Links: Related Structures, BioAssays, BioSystems, Literature, Other Links

All: 120 Rule of 5: 62

Items 1 - 20 of 120 Page 1 of 6 Next

☐ 1: CID: 5816



Related Structures, BioAssays, BioSystems, Literature, Other Links

epinephrine; Adrenalin, adrenaline ...
IUPAC: 4-[(1R)-1-hydroxy-2-(methylamino)ethyl]benzene-1,2-diol
MW: 183.204420 g/mol | MF: C₉H₁₃NO₃
Tested in BioAssays: All: 274, Active: 22, BioActivity Analysis
Vasoconstrictor Agents... more

☐ 2: CID: 838



Related Structures, BioAssays, Literature, Other Links

epinephrine; DL-Adrenaline; Racepinefrine ...
IUPAC: 4-[1-hydroxy-2-(methylamino)ethyl]benzene-1,2-diol
MW: 183.204420 g/mol | MF: C₉H₁₃NO₃
Tested in BioAssays: All: 17, Active: 1, BioActivity Analysis
Vasoconstrictor Agents... more

☐ 3: CID: 247704

Related Structures, BioAssays, Literature, Other Links

Selected Compounds Compound Count

BioActivity Experiments

BioAssays, Active	39
BioAssays, Tested	60
Protein 3D Structures	6
Crystal Structure Of Dipeptide...	1

BioMedical Annotation

Pharmacological Actions	69
Sympathomimetics	37
BioSystems	16

Depositor Category

Biological Properties	120
Chemical Vendors	55
Journal Publishers	47
NIH Molecular Libraries	52

Recent activity

Базы данных химических соединений

Chemical Abstract Service – в регистре содержится 130 млн соединений (2018)

CAS
A division of the American Chemical Society

ACS JOURNALS C&EN CAS

ABOUT CAS OUR EXPERTISE SOLUTIONS PRODUCTS & SERVICES SUPPORT & TRAINING NEWS & EVENTS Site Search

SciFinder
Increase your productivity and achieve faster breakthroughs

STN
Access the most comprehensive scientific, technical, & patent information

55519680
ORGANIC AND INORGANIC
SUBSTANCES
TO DATE

Carbon Bond Formations Win 2010 Nobel Prize with Award-Winning Research in CAS Databases

Check out the early inspiring efforts of these three great chemists.

LEARN MORE

QUICK LINKS
Find all your product resources in SciFinder and STN Support & Training.

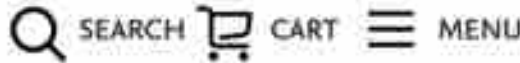
THE RESEARCH EDGE
In the race to discovery, access to information is the critical differentiator. See how to ensure your research is on the cutting edge and stay ahead of the competition.

CAS UPDATES

- CAS adds Costa Rica Registry of Industrial Property as 61st patent authority in its patent coverage
- Access an additional 1.8 million records exclusively

SciFinder®

Базы данных химических соединений



Reaxys

Reaxys retrieves literature, compound properties and chemical reaction data faster than any other solution.

[Get started](#)

Already a Reaxys customer?

If you already have a Reaxys account, please click the link below to sign in.

[> Reaxys sign in](#)

Life is chemistry: Meet the Reaxys team

Our technologists and chemists combine their expertise to create a high-quality information database.

[> Watch the video](#)

Teaching chemistry info literacy

Are you doing enough to prepare future chemists for their careers in research?

[> Take the assessment](#)

[Why choose Reaxys](#)

[How Reaxys works](#)

[Who we serve](#)

[Customer stories](#)

[Learn & support](#)

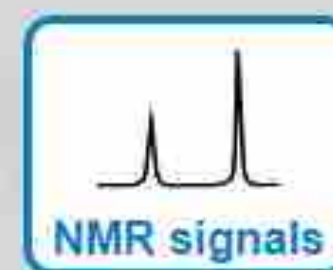
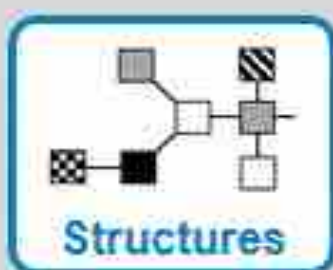
Базы данных углеводов



For 2017:

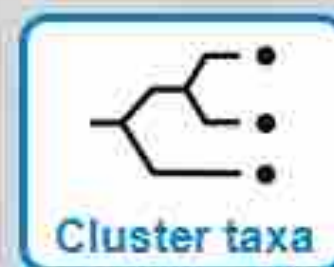
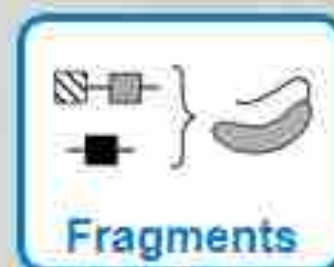
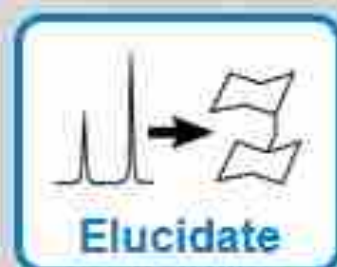
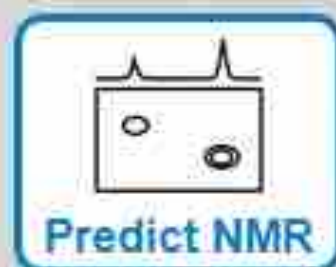
7005 publications for 18924 compounds from 8859 organisms

Database search



Additional operations are available from the [left menu](#). If you don't see it [click here](#)

Useful tools



Структурные базы данных

7009 структур липидов и сходных соединений – не поддерживается в настоящее время ☹

LipidBank

LipidBank

Vitamin A

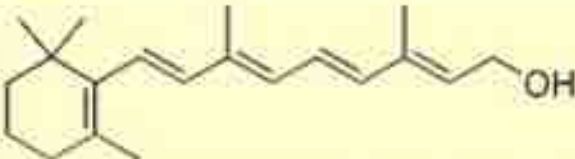
DATA No : VVA0001 **INFORMANT :** Hiroyuki Kagechika

NAME : ((2E,4E,6E,8E)-3,7-Dimethyl-9-(2,6,6-trimethyl-cyclohex-1-enyl)-nona-2,4,6,8-tetraen-1-ol

COMMON NAME : all-trans-retinol

SYMBOL : ROH

FORMULA : C₂₀H₃₀O **MOL.WT (average) :** 286.452



[Download ChemDraw structure data](#)

BIOLOGICAL ACTIVITY

Relative physiological activity for All trans-retinyl acetate is 100%. ([Ref. 0046](#)) New physiological functions of vitamin A have been identified in addition to its role in vision ([Ref. 0005](#)), namely its role in immune defence reducing morbidity of measles, ([Ref. 0006](#)) of respiratory and possibly HIV infections ([Ref. 0007](#)), in gene regulation ([Ref. 0008](#)), in cell differentiation ([Ref. 0009](#)) and morphogenesis ([Ref. 0010/0011](#))

PHYSICAL AND CHEMICAL PROPERTIES

MELTING POINT : 62-64°C([Ref. 0013](#))

План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- **Базы данных геномов и аннотаций**
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

Genome

Genome

Search

[Limits](#) [Advanced](#)

[Help](#)



Genome

This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

Using Genome

[Help](#)

[Browse by Organism](#) **UPDATED**

[Download / FTP](#)

[Download FAQ](#)

[Submit a genome](#)

Custom resources

[Human Genome](#)

[Microbes](#)

[Organelles](#)

[Viruses](#)

Other Resources

[Assembly](#)

[BioProject](#)

[BioSample](#)

[Map Viewer](#)

[Genome Data Viewer](#) **NEW**

Genome Tools

[BLAST the Human Genome](#)

[Microbial Nucleotide BLAST](#)

Genome Annotation and Analysis

[Eukaryotic Genome Annotation](#)

[Prokaryotic Genome Annotation](#)

[PASC \(Pairwise Sequence Comparison\)](#)

External Resources

[GOLD - Genomes Online Database](#)

[Bacteria Genomes at Sanger](#)

[Ensembl](#)



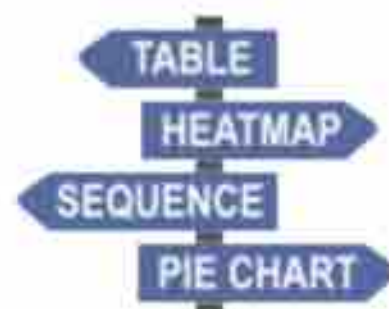
Search: for

e.g. **BRCA2** or **rat 5:62797383-63627669** or **rs699** or **coronary heart disease**

Browse a Genome

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple

Find a Data Display



Not sure how to find the data visualisation you need? With our

What's New in Ensembl Release 91 (December 2017)

- [New Primate Species](#)
- [Update of cat assembly and genebuild to Felis_catus_8.0](#)
- [Mouse: update to Ensembl-Havana GENCODE gene set](#)
- [New and updated probe mapping data for primates](#)
- [Microarray Probe Mapping Update](#)

[Full details](#) | [All web updates](#), [by](#)

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[View](#)[Help](#)[About Us](#)

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

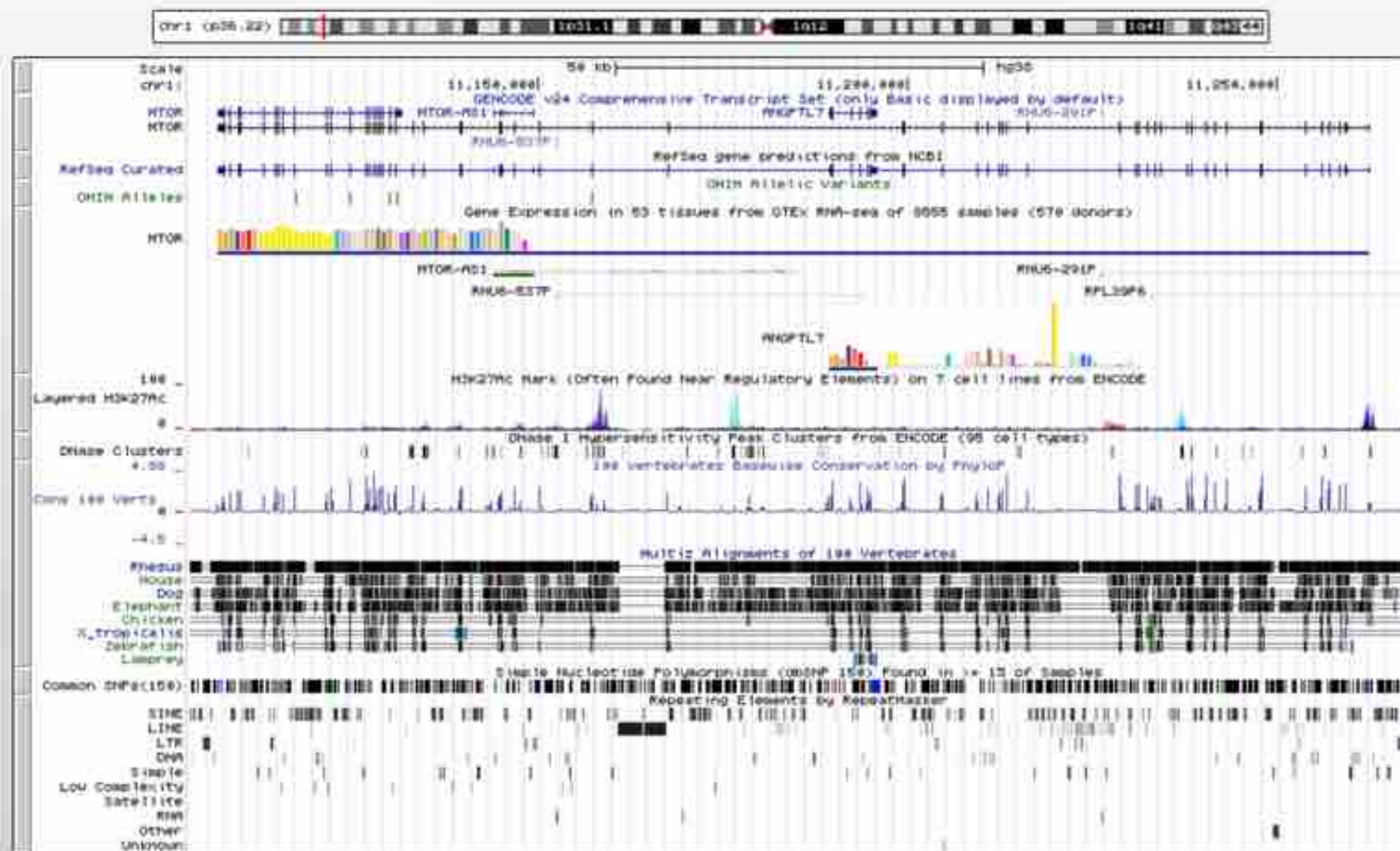
chr1:11,102,837-11,267,747 164,911 bp.

enter position, gene symbol, HGVS or search terms



go

[Request onsite workshops](#)



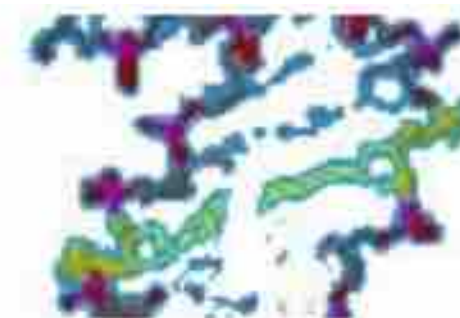
План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- **Базы данных вариаций генома**
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии



dbSNP

Short Genetic Variations

[dbVar](#)[ClinVar](#)[GaP](#)[PubMed](#)[Nucleotide](#)[Protein](#)

Search small variations in dbSNP or large structural variations in dbVar

Search Entrez

dbSNP

for

7

Go

Have a question
about dbSNP? Try
searching the SNP
FAQ Archive!

7

Go

ANNOUNCEMENT

dbSNP and dbVar no longer accept submissions
for non-human organism data. Please read more
[here](#).

GENERAL

[RSS Feed](#)

[Contact Us](#)

[Organism Data](#)

[dbSNP Homepage](#)

Search by IDs on All Assemblies

Note: `refseq` and `ncbi` must be prefixed with "ref" or "ncbi" respectively. For

dbVar

dbVar



Search

[Advanced](#)

[Help](#)



dbVar

dbVar is NCBI's database of human genomic structural variation — insertions, deletions, duplications, inversions, mobile elements, and translocations

Getting Started

[Overview of Structural Variation](#)

Accessing Data

[Structural Variation Data Hub](#)

Other NCBI Resources

[dbSNP](#)

OMIM

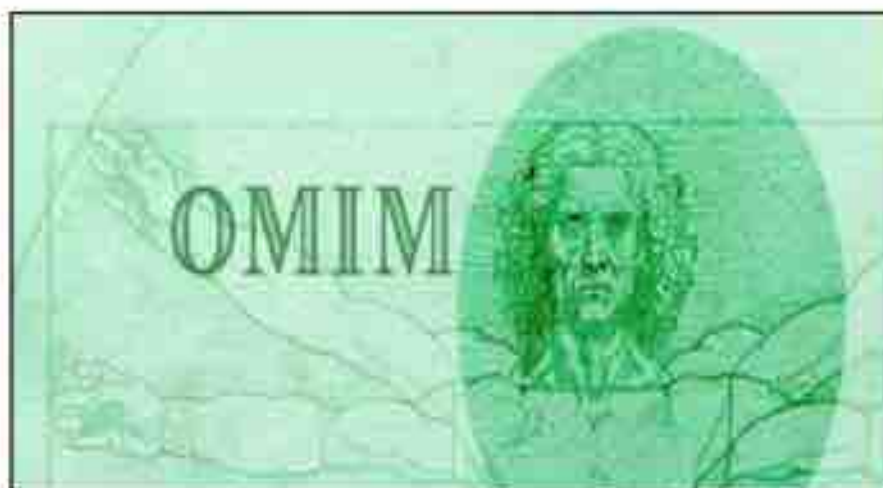
OMIM



Search

[Limits](#) [Advanced](#)

[Help](#)



OMIM

OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada

Using OMIM

[Getting Started](#)

[FAQ](#)

OMIM tools

[OMIM API](#)

Related Resources

[ClinVar](#)

[Gene](#)

[GTR](#)

[MedGen](#)

ClinVar

ClinVar

Search ClinVar for gene symbols, HGVS expressions, conditions, and more

Search

[Advanced](#)

[Help](#)

[Home](#)

[About](#)

[Access](#)

[Help](#)

[Submit](#)

[Statistics](#)

[FTP](#)

ACTGATGGTATGGGGCCAAGAGATATATCT
 CAGGTACGGCTGTCATCACTTAGACCTCAC
 CAGGGCTGGGCATAAAAGTCAGGGCAGAGC
 CCATGGTGCATCTGACTCCTCAGGAGAAGT
 GCAGGTTGGTATCAAGGTTACAAGACAGGT
 GGCACCTGACTCTCTCTGCCTATTGGTCTAT

ClinVar

ClinVar aggregates information about genomic variation and its relationship to human health.

Using ClinVar

[About ClinVar](#)

[Data Dictionary](#)

[Downloads/FTP site](#)

[FAQ](#)

[Contact Us](#)

[RSS feed/What's new?](#)

[Factsheet](#)

Tools

[ACMG Recommendations for Reporting of Incidental Findings](#)

[ClinVar Submission Portal](#)

[Submissions](#)

[Variation Viewer](#)

[Clinical Remapping - Between assemblies and RefSeqGenes](#)

[RefSeqGene/LRG](#)

Related Sites

[ClinGen](#)

[GeneReviews](#)

[GTR](#)

[MedGen](#)

[OMIM](#)

[Variation](#)

План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- **Базы данных геном-фенотип**
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

[Log In](#) to dbGaP

dbGaP Data Download

The management portal to request and download individual level data

Click [here](#) to login to the dbGaP controlled-access portal and to begin a project request. For guidance on the development of a data access request to complete project requests, please see

 [Tips for preparing a successful Data Access Request.](#)

[Who can apply for access?](#)

[How does one apply?](#)

[Why is Access Controlled?](#)

dbGaP Data Browser – View Only

With dbGaP Data Browser approval through the [simplified controlled-access](#) application, users may view the collection “[Compilation of individual-level data from general research use \(GRU\).](#)”

[What is the purpose of the dbGaP Data Browser; why is it useful?](#)

[How does one apply?](#)

[Additional help.](#)



ICGC



Data Portal
Get Cancer Data



Data Access Compliance Office
Apply for Access to Controlled Data

[Contact Us](#)

[Log In](#) | [Create an Account](#)



International
Cancer Genome
Consortium

Enter keywords

Search

Home

[Cancer Genome Projects](#)

[Committees and Working Groups](#)

[Policies and Guidelines](#)

[Media](#)

[Publications](#)

ICGC Cancer Genome Projects

Committed projects to date: [90](#)

Sort by: Project

[Biliary Tract Cancer](#)
Japan

[Biliary Tract Cancer](#)
Singapore

[Bladder Cancer](#)
China

[Bladder Cancer](#)
United States

[Blood Cancer](#)
China

[Blood Cancer](#)
Singapore

ICGC Goal: To obtain a **comprehensive** description of **genomic, transcriptomic and epigenomic changes** in **50 different tumor types and/or subtypes** which are of clinical and societal importance across the globe.

[Read more »](#)

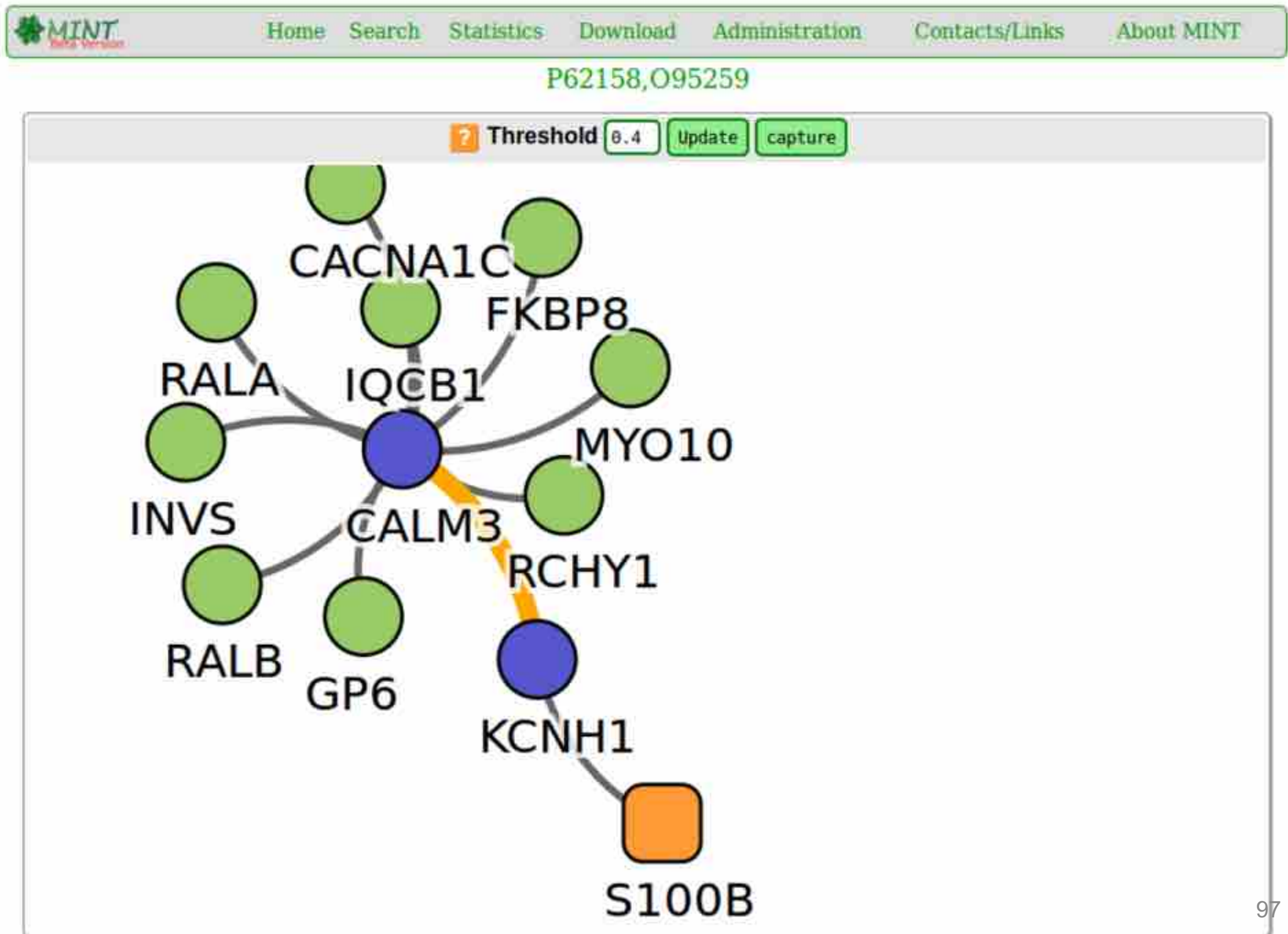
[Launch Data Portal »](#)

[Apply for Access to Controlled Data »](#)

План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- **Базы данных взаимодействий**
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

Базы данных взаимодействий



Базы данных взаимодействий

   Database of potassium channel toxins from scorpion venom Help & FAQs About & Contact									
174 toxins shown									
☐	Organism ▼	Name ▼	Synonyms	UniProt ID	Sequence	PDB	Mass	Pub. Date	Activity ▼
☐	<i>Leiurus quinques...</i>	α-KTx 1.1	ChTX; ChTX-Lq...	P13487	ZFTNVSCCTTSKECWSVCQRLHNTSRGKCMNKKCRGYS	1BAH, 1C...	4295.05	1988	Shaker, Kv1.1, Kv1.2,
☐	<i>Leiurus quinques...</i>	α-KTx 1.2	ChTX-Lq2; ChTx...	P45628	ZFTQESCTASNCQWSICKRLHNTNRGKCMNKKCRGYS	1LIR	4335.08	1989	Kv1, KCa1.1
☐	<i>Mesobuthus tam...</i>	α-KTx 1.3	IbTx; Iberiotoxin	P24663	ZFTDVCDSVSKECWSVCKDLFGVDRGKCMGKKCRGYS		4230.02	1990	KCa1.1
☐	<i>Mesobuthus mar...</i>	α-KTx 1.5	BmTX1; Neuroto...	Q9NII6	ZFTDVKCTGSKQCPVCKQMFQKPNKGCMNGKCRGYS	1BKG	4169.04	1997	Kv1, Kv1.3, KCa1.1
☐	<i>Mesobuthus mar...</i>	α-KTx 1.6	BmTX2; Neuroto...	Q9NII5	ZFTNVSCSASSQCPVCKKLFQTYRGKCMNSKCRGYS	2BMT	4178.97	1997	Kv1, Kv1.3, Kv1.1.1, K...
☐	<i>Parabuthus trans...</i>	α-KTx 1.10	PBTx3; Parabuto...	P83112	EVDMRCKSSKECLVKCKQATGRPNGKCMNRKCKCYPR		4274.23	2002	Kv1.1, Kv1.2, Kv1.3, K...
☐	<i>Centruroides nox...</i>	α-KTx 1.11	SloTx; Slo toxin	P0C182	TFIDVDCCTVSKECWAPCKAAFGVDRGKCMGKKCKCYV		4085.98	2001	KCa1.1
☐	<i>Mesobuthus eup...</i>	α-KTx 1.16	Toxin MeKTx11-1	C0HJQ8	ZFTDVKCTGTQCPVCKKMFGRPNGKCMNGKCRGYP		4221.16	2015	Kv1.1
☐	<i>Mesobuthus eup...</i>	α-KTx 1.17	Toxin MeKTx11-3	C0HJQ7	ZFTDVKCTVTQCPVCKKMFGRPNGKCMNGKCRGYS		4253.21	2015	Kv1.1
☐	<i>Centruroides nox...</i>	α-KTx 2.1	NTx; Noxiustoxin...	P08815	TIINVKCTSPKQCSKPCKELYGSSAGAKCMNGKCKCYNN-NH ₂	1SXM	4195.06	1982	Shaker, Kv1.1, Kv1.2,
☐	<i>Centruroides mar...</i>	α-KTx 2.2	MgTx; Margatoxin	P40755	TIINVKCTSPKQCLPPCKAQFGQSAGAKCMNGKCKCYPH	1MTX	4179.11	1993	Shaker, Kv1.1, Kv1.1,
☐	<i>Centruroides lim...</i>	α-KTx 2.3	ClITx1; Toxin I, T...	P45629	ITINVKCTSPQCLRPCKDRFGQHAGGKCKINGKCKCYPH		4191.06	1994	Kv1
☐	<i>Centruroides nox...</i>	α-KTx 2.4	NTx-2; NTx2; No...	Q9TXD1	TIINEKCFATSCQWTPCKKAIGSLQSKCMNGKCKCYNG		4183.05	1996	Kv1
☐	<i>Centruroides lim...</i>	α-KTx 2.5	HgTX1; Hongoto...	P59847	TVIDVKCTSPKQCLPPCKAQFGIRAGAKCMNGKCKCYPH	1HLY	4220.20	1998	Kv1.1, Kv1.2, Kv1.3, K...
☐	<i>Centruroides lim...</i>	α-KTx 2.7	ClITx2; Toxin II, T...	P45630	TVIDVKCTSPKQCLPPCKEYGRHAGAKCMNGKCKCY		3905.82	1994	Kv1

Базы данных химических соединений

 DRUGBANK



WHAT ARE YOU LOOKING FOR?

Tylenol



Drugs



Targets



Pathways



Indications

 DRUGBANK

План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- **Базы данных сигнальных/метаболических путей**
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions and relations

[Menu](#) [PATHWAY](#) [BRITE](#) [MODULE](#) [KO](#) [GENES](#) [LIGAND](#) [NETWORK](#) [DISEASE](#) [DRUG](#) [DBGET](#)

Select prefix

Enter keywords

[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, reaction and relation 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

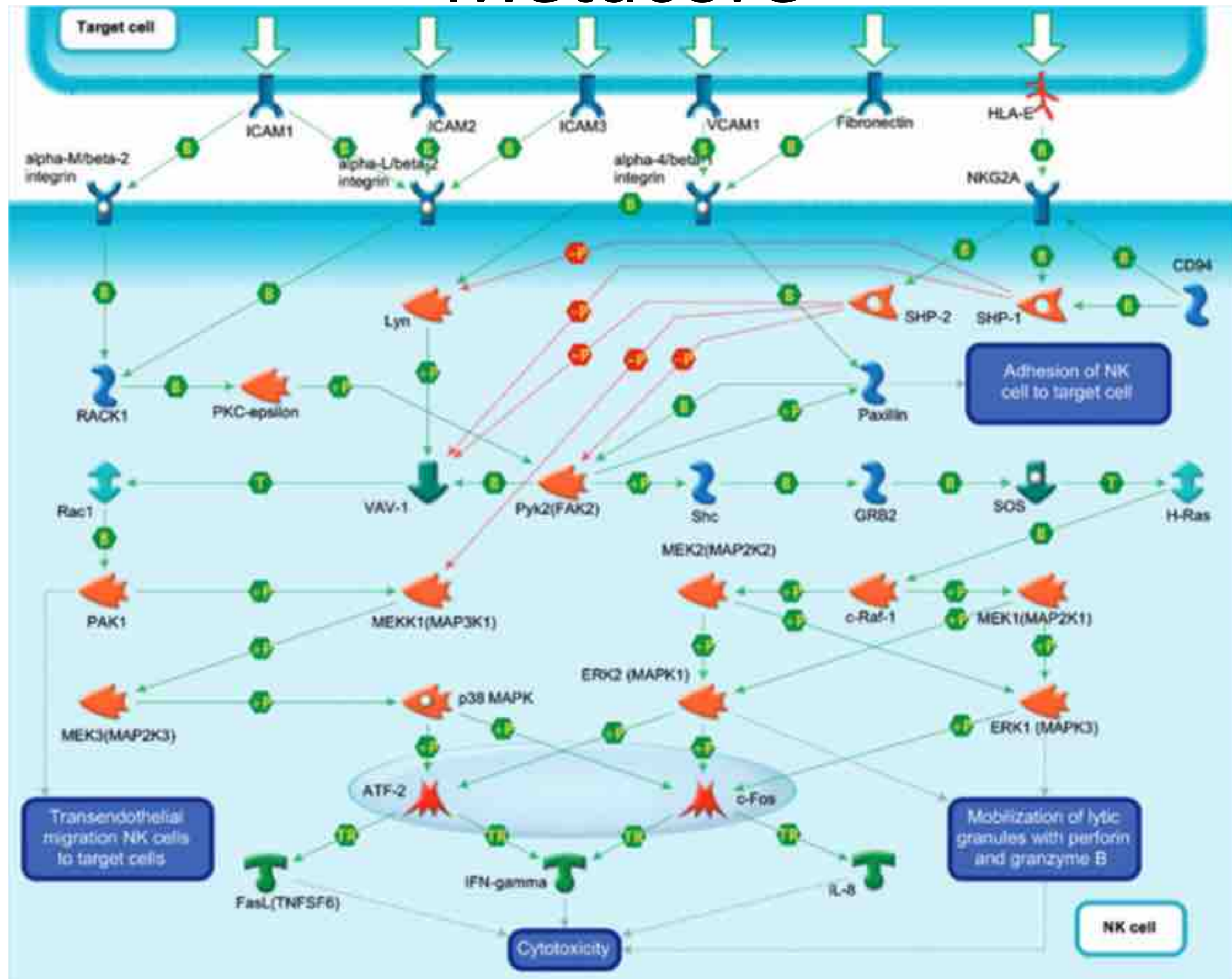
5. Organismal Systems

6. Human Diseases

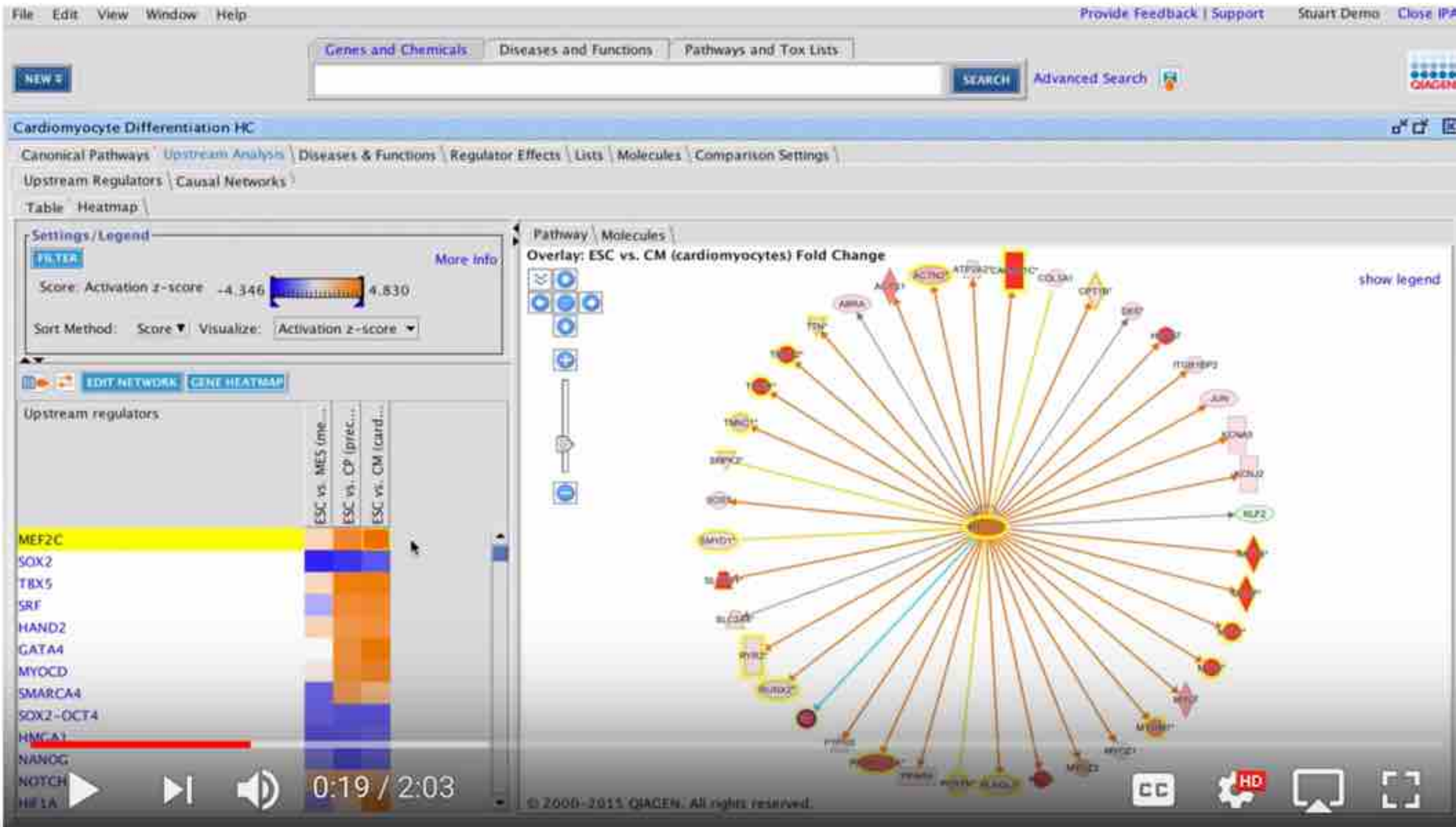
7. Drug Development

KEGG PATHWAY is a reference database for [Pathway Mapping](#).

Metacore



Ingenuity Pathway Analysis



Pathway Studio®



План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- **Базы данных секвенирования**
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

SRA

SRA



[Advanced](#)

Search

[Help](#)



SRA

Sequence Read Archive (SRA) makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®.

Getting Started

[How to Submit](#)

[Log in to SRA \(for updating and troubleshooting submissions\)](#)

[Log in to Submission Portal \(for submitting sequence data\)](#)

Tools and Software

[Download SRA Toolkit](#)

[SRA Toolkit Documentation](#)

[SRA-BLAST](#)

[SRA Run Browser](#)

Related Resources

[Submission Portal](#)

[Trace Archive](#)

[dbGaP Home](#)

[BioProject](#)

План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- **Базы данных клинических исследований и лекарств**
- Базы данных по экспрессии генов/гистологии
- Базы данных по таксономии

ClinicalTrials.gov is a database of privately and publicly funded clinical studies conducted around the world.

Pharmaprojects: track pharma R&D

End-to-end tracking of the global pharma R&D pipeline from bench to patient, including company development trends, global development status, and therapeutic class status.

The industry's most trusted drug development database, Citeline's Pharmaprojects has been covering pharma R&D across global markets for 35+ years. With 68,000+ drug profiles including 15,000 drugs in active development, it's the go-to resource for preclinical, clinical, and pipeline coverage, and lifecycle management tracking.

[Request a free demo >](#)[Product login](#)

-  Curated by experts
-  Drug R&D landscape
-  Historical trends



План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- **Базы данных по экспрессии генов/гистологии**
- Базы данные по таксономии

Базы данных экспрессии генов



Expression Atlas

Examples: ASPM, REACT_200624, ENSMUSG00000021789, zinc finger

[Search](#)

[Home](#) [Release notes](#) [FAQs](#) [Download](#) [Help](#) [About](#) [Feedback](#)

Strand-specific RNA-seq of nine chicken tissues

Organism(s): *Gallus gallus*



Gene query ?

KCNA1 x |

☒ Exact match

Organism part ?

(any organism parts)

☐ Specific ?

Expression level cutoff ?

0.4

[Apply](#)
[Reset](#)



Showing 1 of 1 genes found:



Display levels

	brain	colon	heart	kidney	liver	lung	skeletal muscle	spleen	testis
Gene									
KCNA1									



Базы данных экспрессии генов

 [Resources](#)  [How To](#) 

[GEO Home](#) | [Documentation](#)  | [Query & Browse](#)  | [Email GEO](#)

[Sign in to NCBI](#)


Gene Expression Omnibus

Getting Started

- [Overview](#)
- [FAQ](#)
- [About GEO DataSets](#)
- [About GEO Profiles](#)

Tools

- [Search for Studies at GEO DataSets](#)
- [Search for Gene Expression at GEO Profiles](#)
- [Search GEO Documentation](#)

Browse Content

Repository Browser

DataSets:	4348
Series: 	94968
Platforms:	18150

THE HUMAN PROTEIN ATLAS

≡ MENU

HELP

NEWS

SEARCH¹

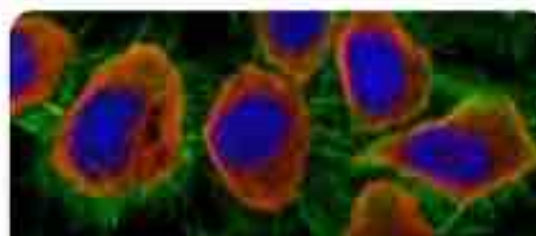
Search

[Fields »](#)

e.g. [RBM3](#), [insulin](#), [CD36](#)



TISSUE ATLAS



CELL ATLAS



PATHOLOGY ATLAS

План

- Библиографические/реферативные базы данных литературных источников (статьи, тезисы, патенты, материалы конференций и т.д.)
- Базы данных последовательностей ДНК
- Базы данных последовательностей белков
- Базы данных 3D структур
- Базы данных хим. соединений
- Базы данных геномов и аннотаций
- Базы данных вариаций генома
- Базы данных геном-фенотип
- Базы данных взаимодействий
- Базы данных сигнальных путей
- Базы данных секвенирования
- Базы данных заболеваний и медицинской информации
- Базы данных по экспрессии генов/гистологии
- **Базы данных по таксономии**

Таксономические базы данных

Taxonomy Browser – знаменитая таксономическая БД, имеющая иерархическую структуру, основанную на анализе последовательностей и призванная упорядочить классификацию организмов, для которых известна хотя бы одна последовательность ДНК или белка.

NCBI Taxonomy Browser

PubMed Entrez BLAST OMIM Taxonomy Structure

Search for As ☒ lock Go Clear

The NCBI Taxonomy Homepage

These are direct links to some of the organisms commonly used in molecular research projects:

Arabidopsis thaliana	Escherichia coli	Pneumocystis carinii
Bos taurus	Hepatitis C virus	Rattus norvegicus
Caenorhabditis elegans	Homo sapiens	Saccharomyces cerevisiae
Chlamydomonas reinhardtii	Mus musculus	Schizosaccharomyces pombe
Danio rerio (zebrafish)	Mycoplasma pneumoniae	Takifugu rubripes
Dictyostelium discoideum	Oryza sativa	Xenopus laevis
Drosophila melanogaster	Plasmodium falciparum	Zea mays

Taxonomy browser
Archaea
Bacteria
Eukaryota
Viroids
Viruses

[Taxonomy common tree](#)

[Taxonomy information](#)

[Taxonomy resources](#)

[Taxonomic advisors](#)

[Genetic codes](#)

Видовые базы данных

Содержат таксономическую, библиографическую, географическую, визуальную и прочую информацию



143,152 species and infraspecific names are in the database, 18,634 images, 54,056 bibliographic items, 314,920 distributional records.

Haematococcus pluvialis Flotow

Publication details

Haematococcus pluvialis Flotow 1844: 415, 537, pls XXIV, XXV

Published in: Flotow, J. von (1844). Beobachtungen über *Haematococcus pluvialis*. Verhandlungen der Kaiserlichen Leopoldinisch-Carolinischen Deutschen Akademie der Naturforscher 12(Abt. 2): 413-606, 3 pls.

 Download PDF

Type species

This is the type species (lectotype) of the genus *Haematococcus*.

Status of name

This name is of an entity that is currently accepted taxonomically.



 *Haematococcus pluvialis* Flotow India

Classification:

Empire Eukaryota
Kingdom Plantae
Subkingdom Viridiplantae
Infrakingdom Chlorophyta
Phylum Chlorophyta
Subphylum Chlorophytina
Class Chlorophyceae
Order Chlamydomonadales
Family Haematococcaceae
Genus *Haematococcus*

Taxonomy

References

Submit Feedback

Submit Reference

Links

Видовые базы данных

<https://plant.depo.msu.ru>



ДЕПОЗИТАРИЙ
ЖИВЫХ СИСТЕМ
«НОЕВ КОВЧЕГ»

[Микроорганизмы и грибы](#) [Растения](#)

[RU](#) [EN](#) [Справка](#)

[О системе](#) [Коллекции](#) [Контакты](#) [Ссылки](#)

 [Вход в систему](#)

Сейчас в базе данных:

 Объектов: 501892

 Изображений: 501892

 Видов: 9141

[Поиск](#)

[Расширенный поиск](#)

[Таксономический словарь](#)

[Коллекции](#)

Национальный банк-депозитарий живых систем

Проект Московского университета "Ноев ковчег" посвящен созданию многофункционального сетевого хранилища биологического материала.

Планируется работа с материалом всех возможных типов - от отдельных биологических молекул до целых живых организмов.

Создание депозитария позволит сохранить биоразнообразие нашей планеты и создать новые способы полезного использования биологического материала.

Фото дня



Geum coccineum

Организм недели



Jurinea compressa

2015-2017

[Подавать образец](#) 