

Введение в биоинформатику

Биологические базы данных

Алексей Александрович Лагунин,

(зав. кафедрой биоинформатики МБФ, д.б.н., профессор РАН)

Борис Владимирович Шилов

(доцент, к.м.н.)

Сергей Михайлович Иванов

(преподаватель)

Комнаты 113, 117



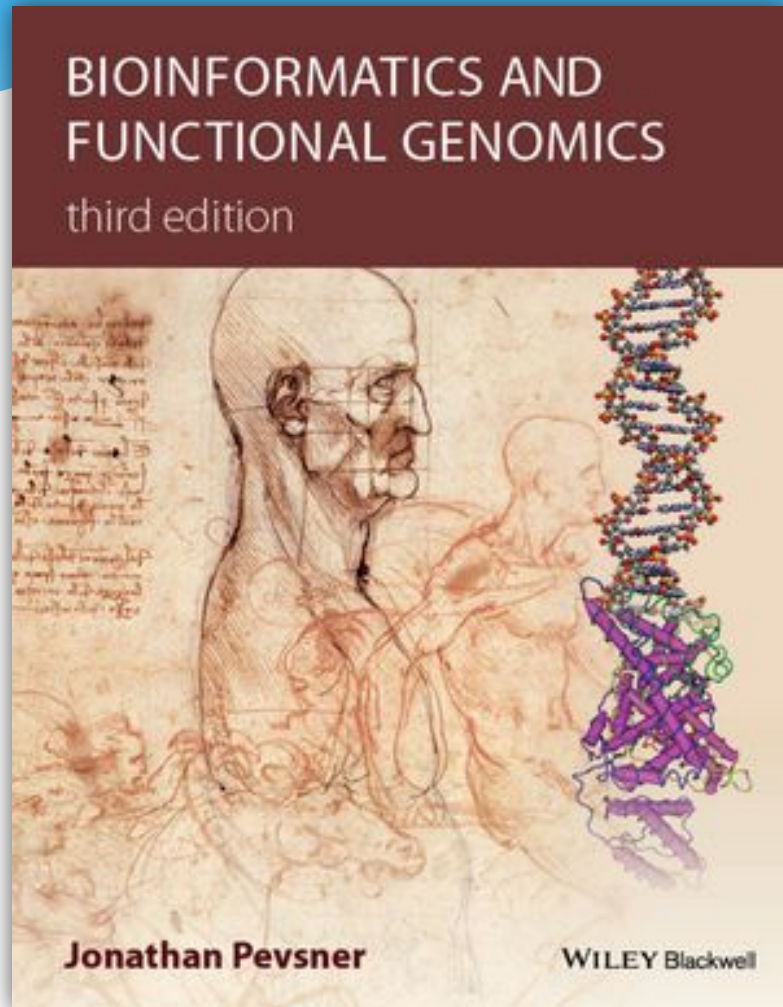
Кафедра биоинформатики МБФ

- * Создана в 2014 году для преподавания биоинформатики и связанных с ней дисциплин студентам Университета и является первой кафедрой биоинформатики в российских медицинских вузах.
- * Направление научных интересов кафедры лежит в области объединения подходов и методов системной биологии, хемо- и биоинформатики в медико-биологических исследованиях и компьютерном конструировании лекарств.
- * Научная база: отдел биоинформатики ФГБУ НИИ биомедицинской химии им. В.Н. Ореховича (<http://www.ibmc.msk.ru/ru/departments/60>)

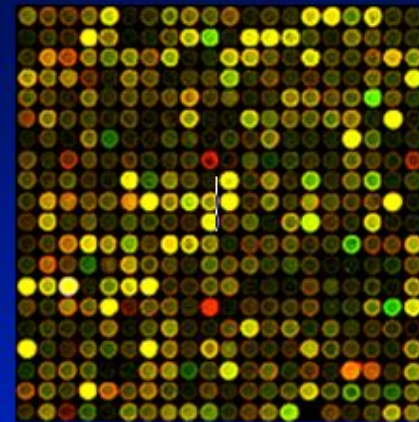
Научно-практические направления работы

- * - Разработка и использование методов анализа «структура-активность» в компьютерной токсикологии и конструировании лекарств.
- * - Использование методов системной биологии для поиска лекарственных мишеней.
- * - Использование дескрипторов нуклеотидных и аминокислотных остатков для анализа сходства и функциональных особенностей биологических последовательностей.
- * - Предсказание лекарственно-индуцированной генной экспрессии и ее использование для анализа возможных фармакологических эффектов лекарств.

Основа курса



Introduction to Bioinformatics



Monday, November 19, 2012
Jonathan Pevsner
pevsner@kennedykrieger.org
Bioinformatics
M.E:800.707

Цели курса:

- Определить понятия и термины используемые в биоинформатике
- Научиться взаимодействовать со специализированными базами данных и интернет-ресурсами
- Изучить основные принципы и методы биоинформатики
- Получить практические навыки работы с программами, применяемыми в биоинформатике

Определение дисциплины

- * <http://www.bisti.nih.gov/docs/CompuBioDef.pdf>
- * 1. Биоинформатика применяет принципы информационных наук и технологий, чтобы сделать огромные, разнообразные и сложные данные наук о жизни более понятными и полезными.
- * 2. Исследование, разработка или применение вычислительных средств и подходов для расширения использования биологических, медицинских и поведенческих данных, а так же для пополнения, хранения, организации, анализа или визуализации таких данных.

Биоинформатика

- * Биоинформатика — совокупность методов и подходов, включающих в себя:
 - * математические методы компьютерного анализа в сравнительной геномике (геномная биоинформатика)
 - * разработку алгоритмов и программ для предсказания пространственной структуры белков (структурная биоинформатика)
 - * исследование стратегий, соответствующих вычислительных методологий, а также общее управление информационной сложности биологических систем
- * В биоинформатике используются методы прикладной математики, статистики и информатики. Биоинформатика используется в биохимии, биофизике, медицине, генетике, экологии и в других областях.

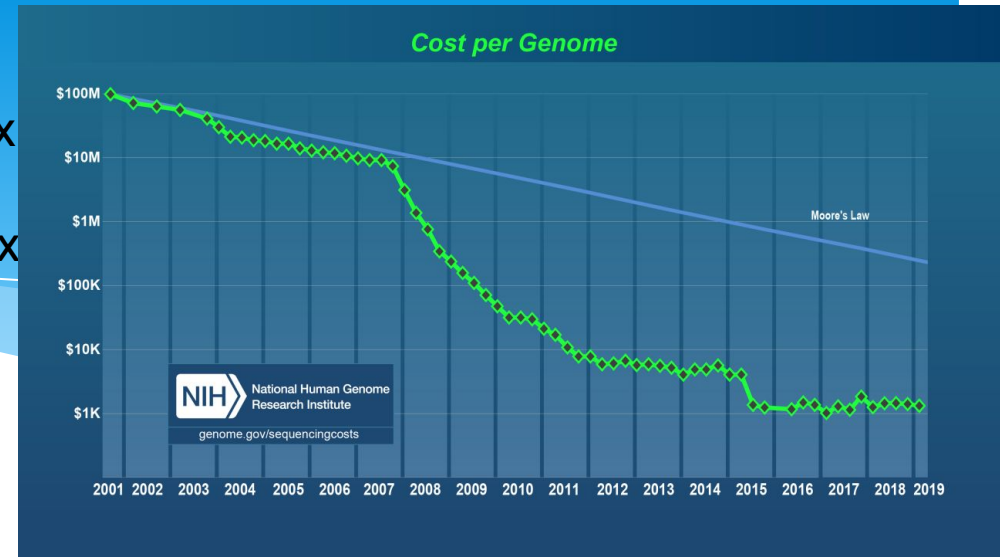
Развитие биоинформатики

Снижение стоимости получения данных

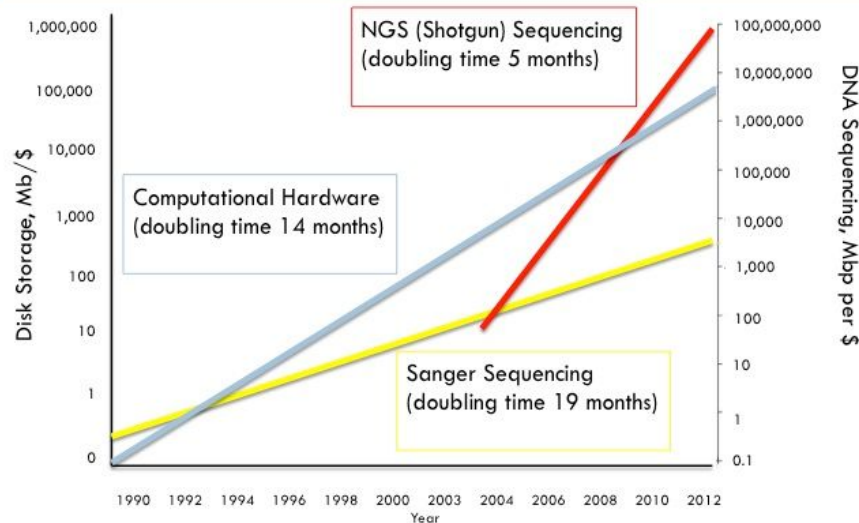
Увеличение скорости получения данных

Увеличение количества данных

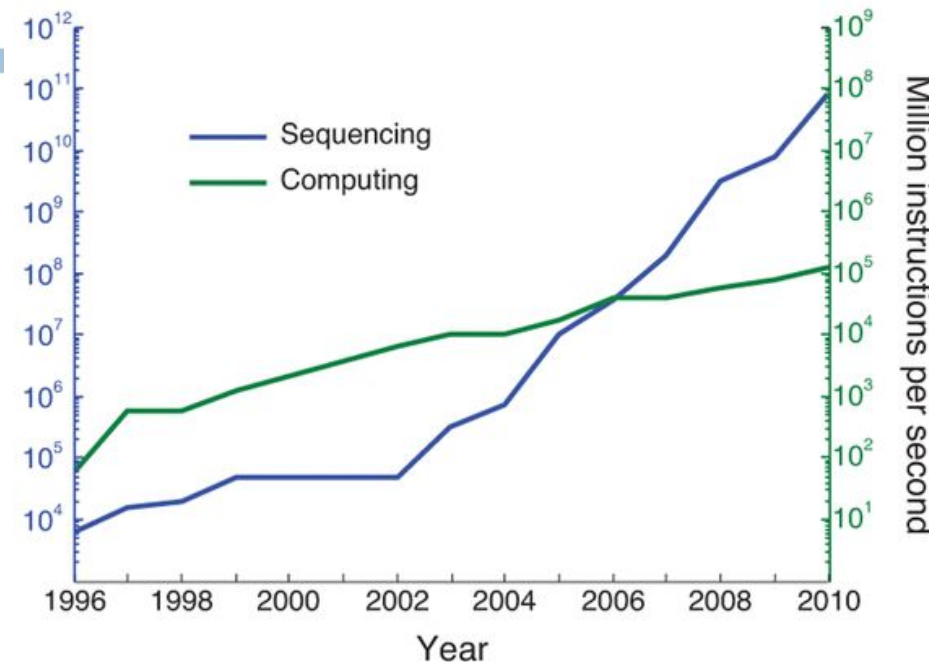
Улучшение доступности данных



The era of big data in biology



Stein, Genome Biology, 2010



Цикл накопления знаний



Цикл накопления знаний

Исследования

Анализ
геномов

База
данных

Данные

10^{12} пар оснований

>500000 биологических видов

>1000 полностью секвенированных геномов

...

Интеграция
знаний

Знания
**Область
применения
биоинформатик
и**

Узкое
место

Интеграция
данных

Анализ данных

Гипотетический биоинформатический процесс

Новый вирус!!!

Секвенирование ДНК (РНК) вируса

Скрининг генома по БД

Поиск белка-мишени (репликация, сборка и т.п.)

Моделирование 3D структуры белка (гомологичное, ab initio)

Поиск сайтов белка

Поиск или создание молекулы для блокирования сайта

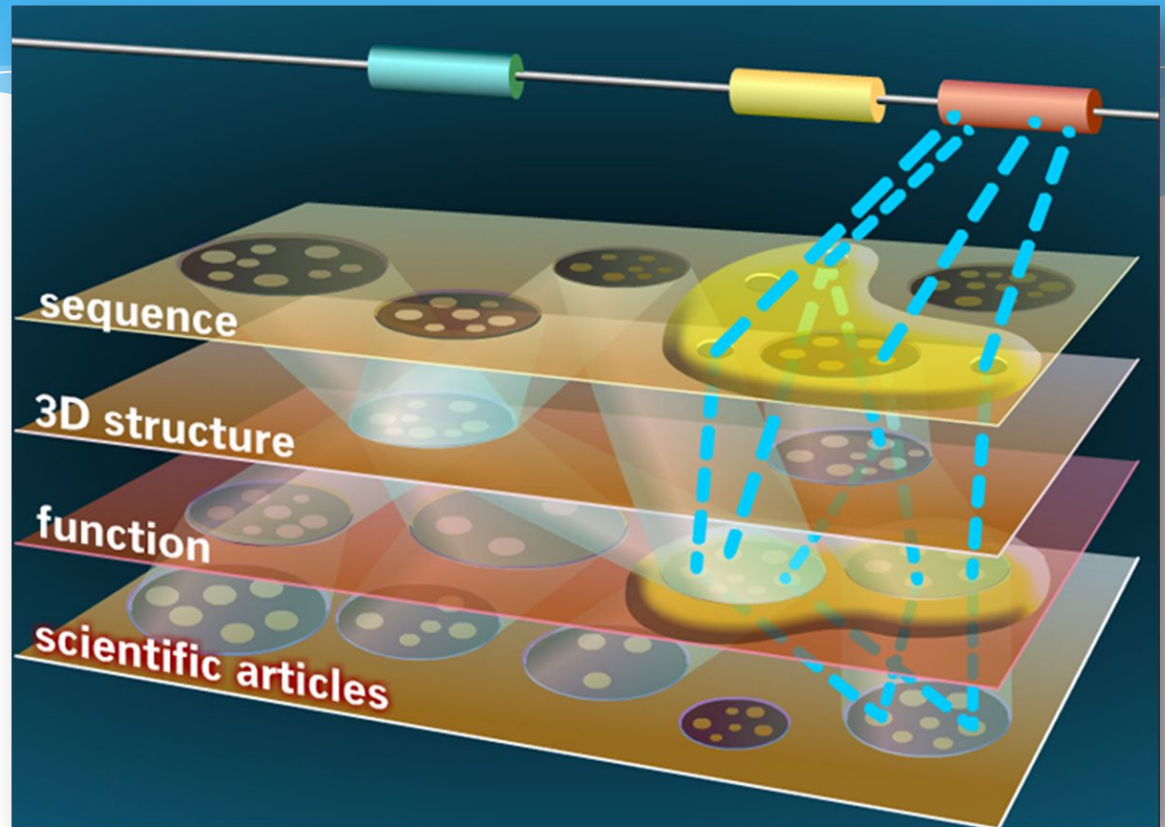
Фармакологические
исследования

Три уровня биоинформатики

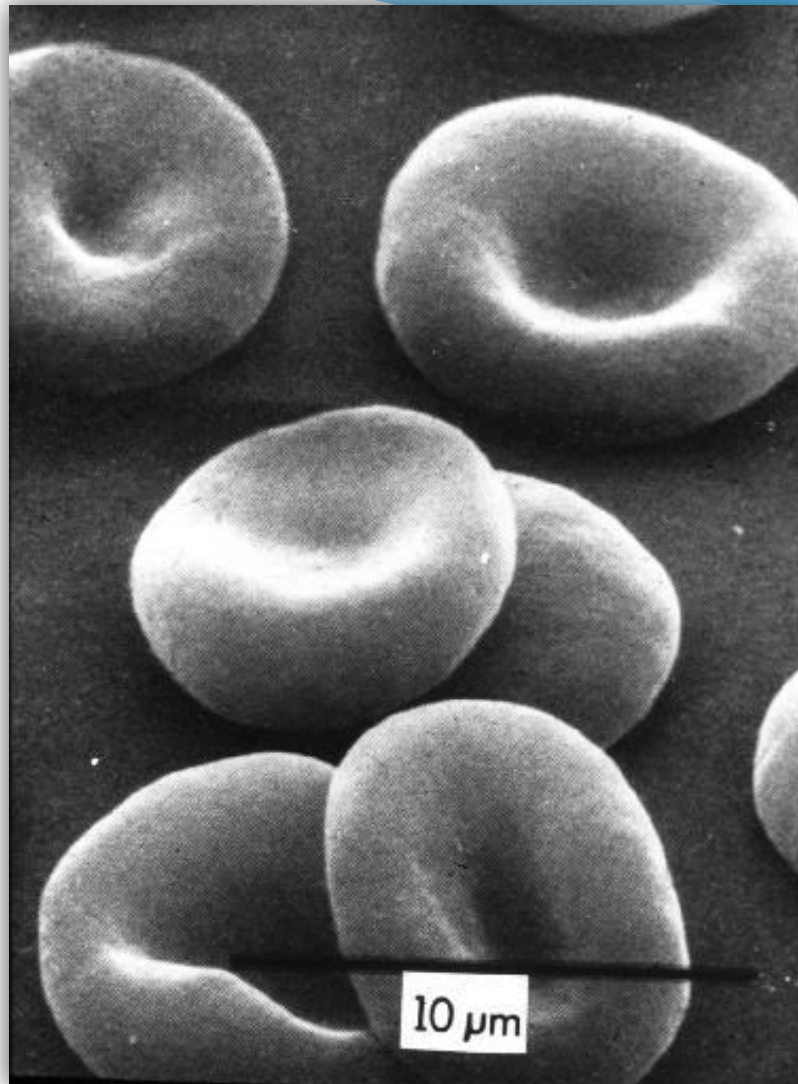
- * Клетка

- * Организм

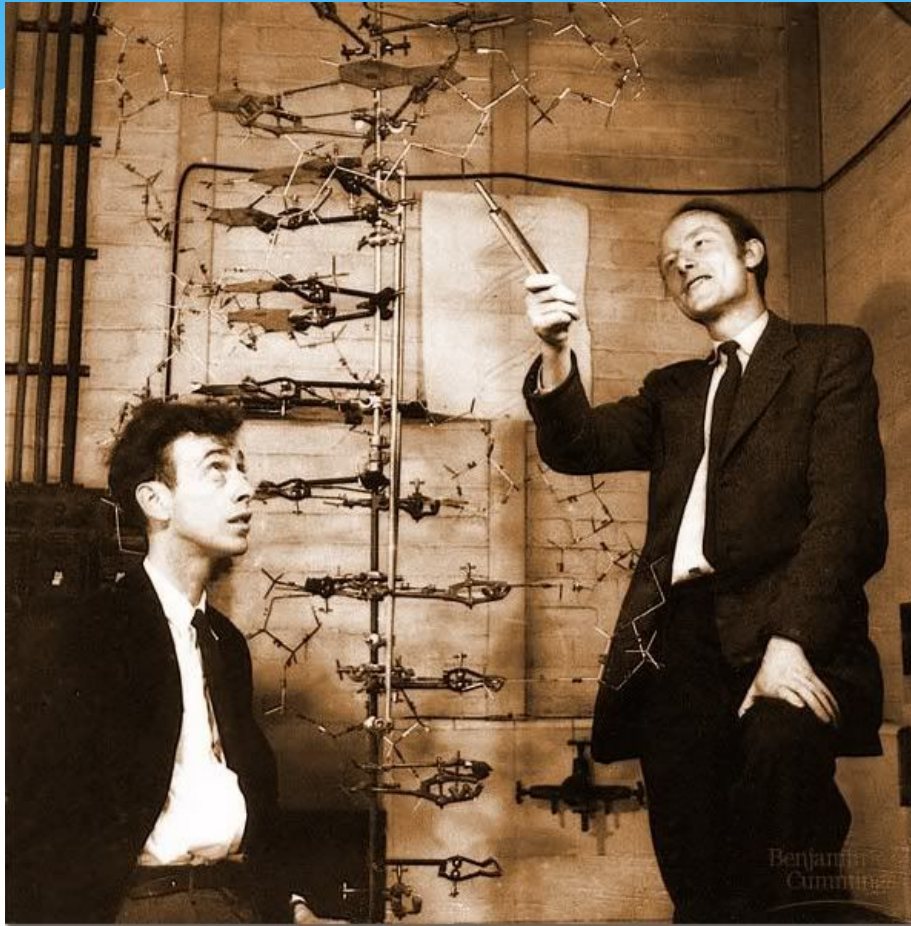
- * Дерево жизни



Первый уровень: Клетка



Джеймс Уотсон, Френсис Крик

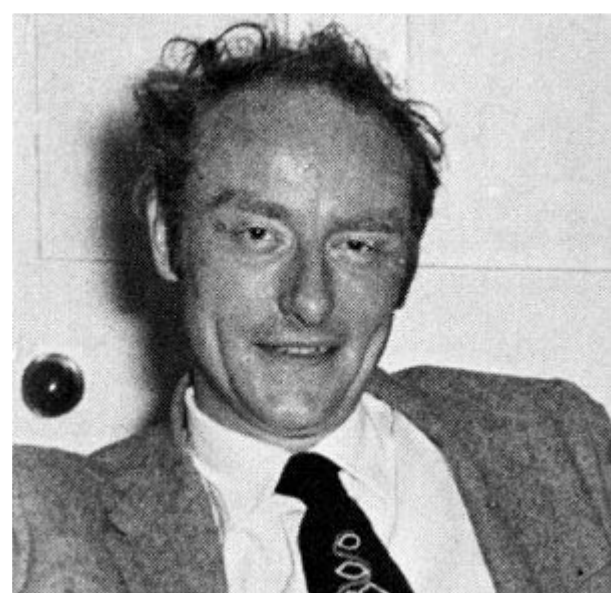


1953 г., USA

РНИМУ



2003 г. Miami, USA

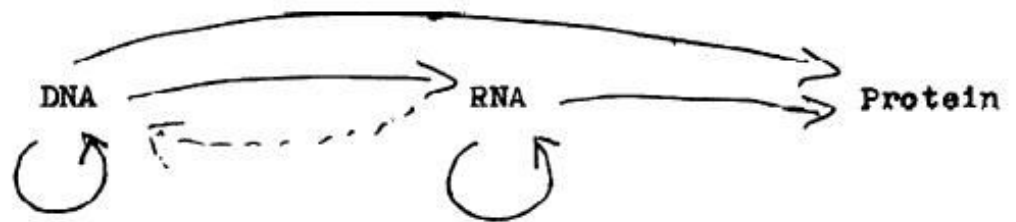


Ideas on Protein Synthesis (Oct. 1956)

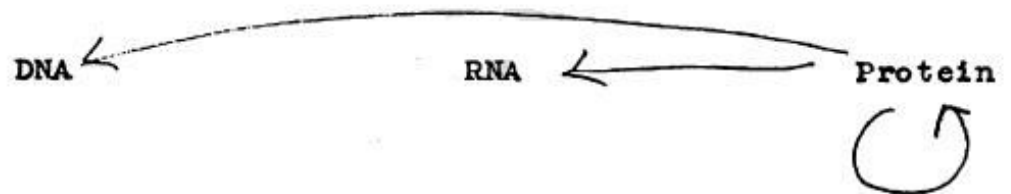
The Doctrine of the Triad.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it.

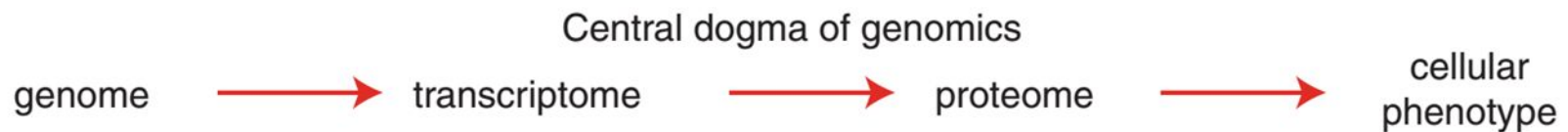
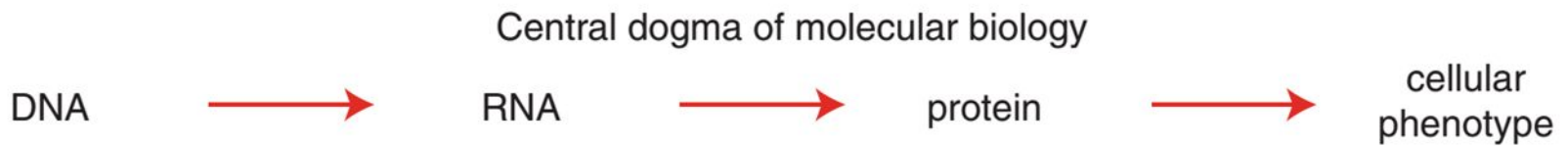
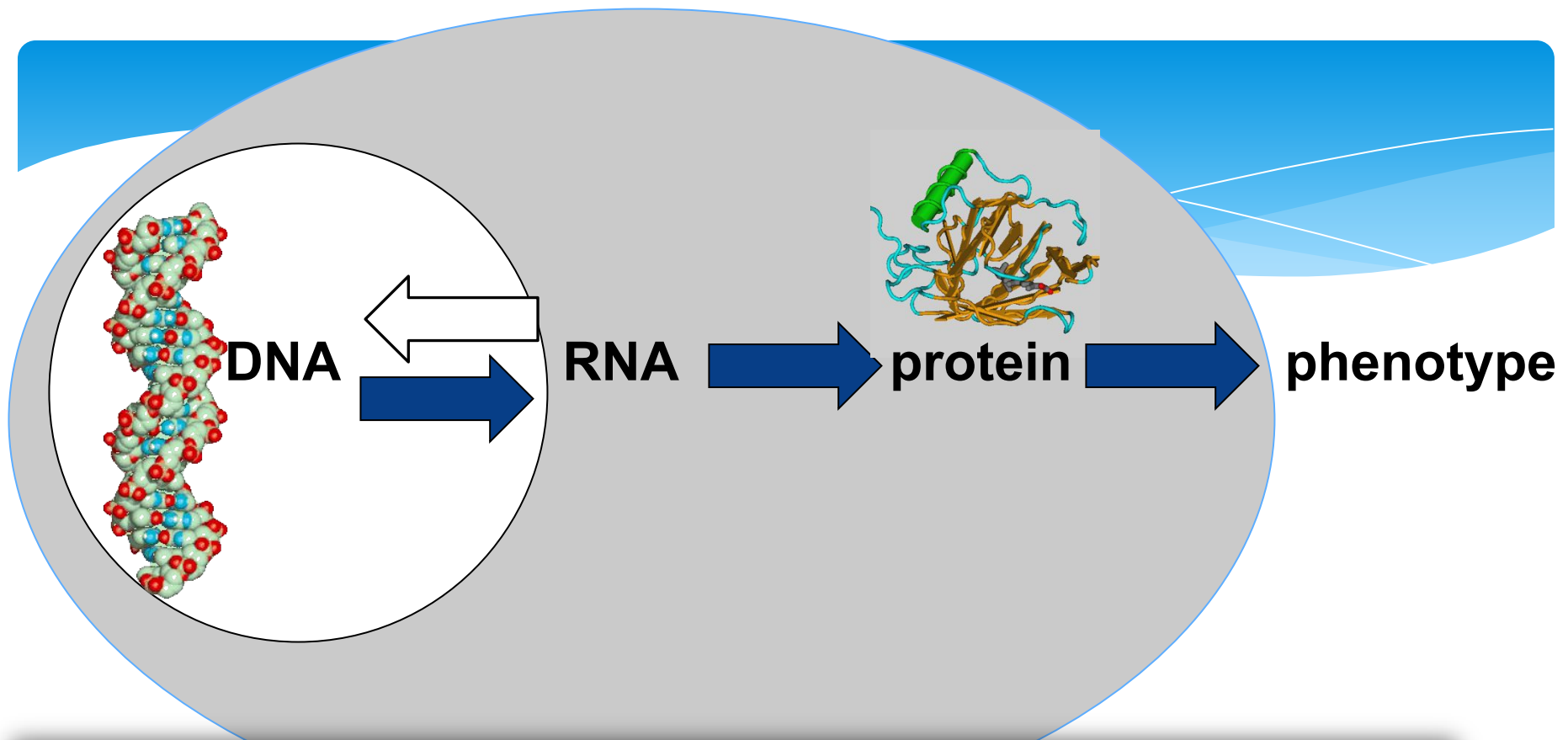
That is, we may be able to have

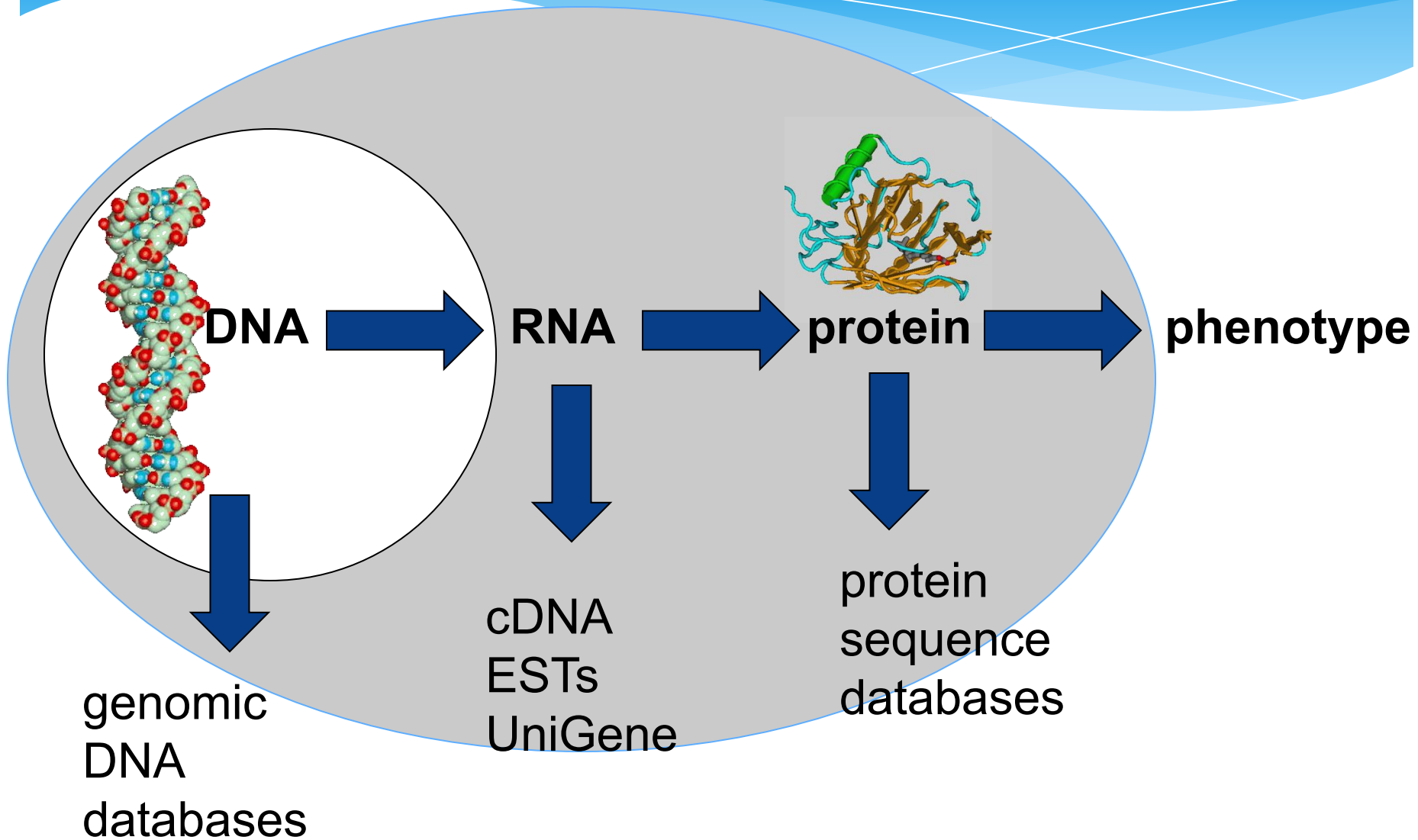


but never



where the arrows¹⁵ show the transfer of information.

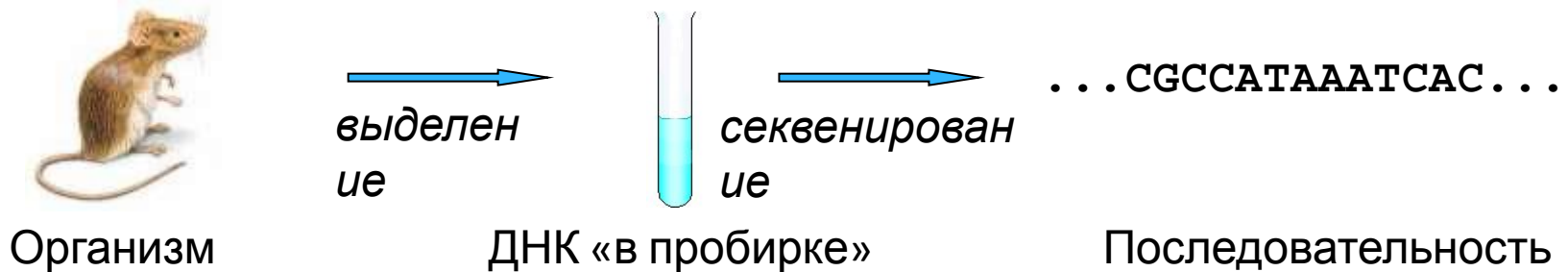




Секвенирование ДНК

Секвенирование — определение аминокислотной (белки) или нуклеотидной (ДНК / РНК) последовательности.

В конце 1970-х годов был открыт относительно быстрый метод расшифровки последовательности оснований в ДНК (Sanger)



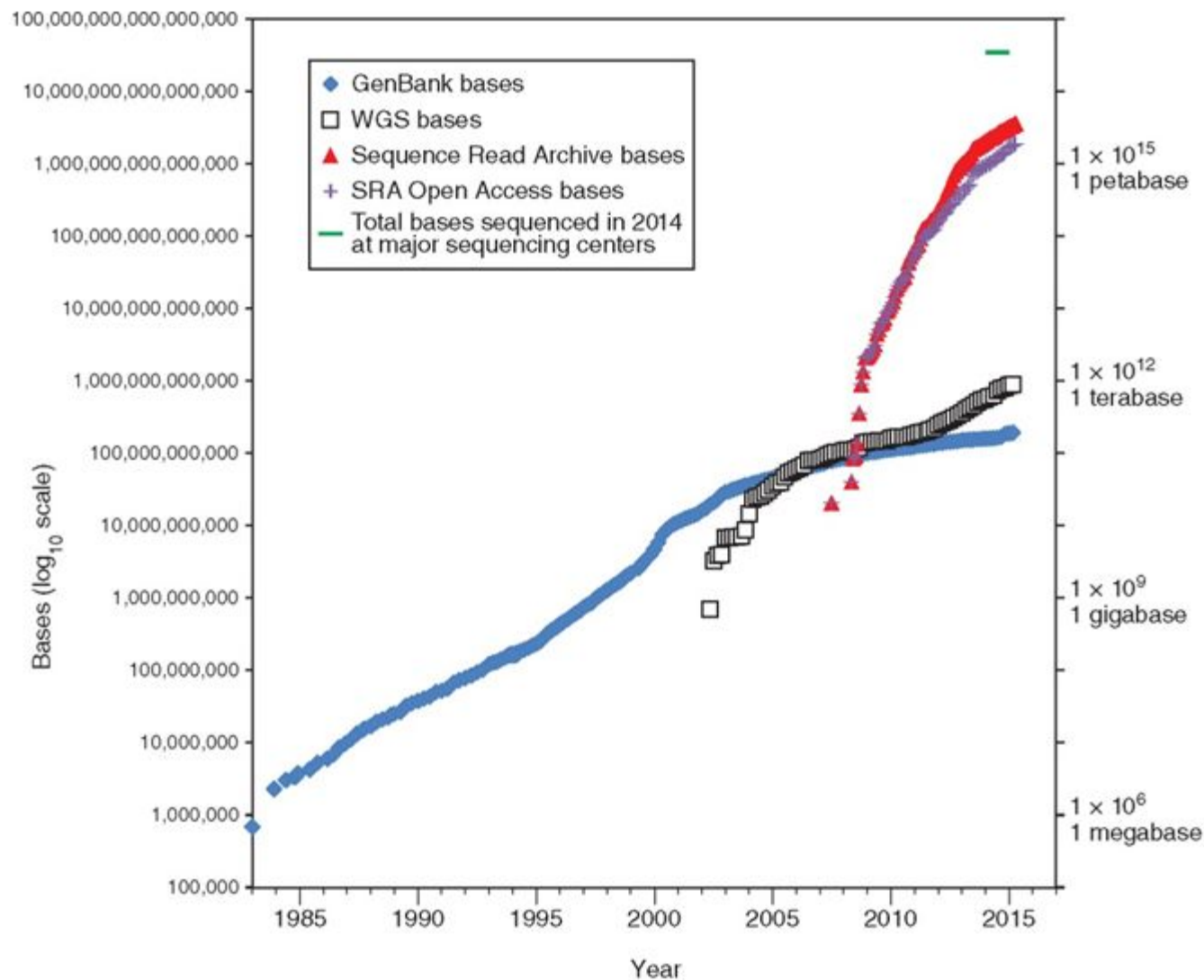
gatcctccatatacaacggtatctccacctcaggttttagatctcaacaacggaaccattg
ccgacatgagacagtttaggtatcgtcgagagttacaagctaaaacgagcagtagtcagct
ctgcatctgaagccgctgaagttctactaagggtggataacatcatccgtgcaagaccaa
gaaccgccaatagacaacatatgtaacatatcttaggatatacctcgaaaataataaaccg
ccacactgtcattattataattagaaacagaacgcaaaaattatccactatataattcaa
agacgcgaaaaaaaaaagaacaacgcgctcatagaacttttggcaattcgcgctcaciaataa
atcttggcaacttatgtttcctcttcgagcagtagctcgagccctgtctcaagaatgtaat
aataccatcgtaggtatggttaaagatagcatctccacaacctcaaagctccttgccga
gagtcgcccctcctttgtcgagtaatcttctacttttcatatgagaacttatcttcttattc
ttactctcacatcctgtagtgattgacactgcaacagccaccatcactagaagaacaga
acaattactttaatagaaaaattatatcttcctcgaaacgatttcctgcttccaacatcta
cgtatatcaagaagcattcacttaccatgacacagcttcagatttcattattgctgacag
ctactatatcactactccatctagtagtggccacgcccctatgaggcatatcctatcggaa
aacaataccccccagtgggcaagagtcfaatcgtttacatttcaaatttccaatgata
cctataaatcgctctgtagacaagacagctcaaataacatacaattgcttcgacttaccga
gctggcttttcgcttgactctagttctagaacgttctcaggtgaaccttcttctgacttac
tatctgatgcgaacaccacgcttgatcttcaatgtaatactcgaggggtacggactctgccg
acagcacgtctttgaacaatacataccaatttggtgttacaacacgctccatccatctcgc
tatcgtcagatttcaatctattggcgcttggttaaaaaactatgggttatactaacggcaaaa
acgctctgaaactagatcctaatagaagtcttcaacgtgacttttgaccgttcaatgttca
ctaacgaagaatccattgtgtcgtattacggacgcttctcagttgtataatgcgccgttac
ccaattggctgttcttcgattctggcgagttgaagtttactgggacggcaccggtgataa
actcggcgattgctccagaaacaagctacagttttgtcatcatcgctacagacattgaag
gattttctgccgttgaggtagaattcgaattagtcatcggggctcaccagttaactacct
ctattcaaaaatagtttgataatcaacgttactgacacaggtaacgtttcatatgacttac
ctctaaactatgtttatctcgatgacgatcctattttcttctgataaattgggttctataa

Шкала ДНК пар оснований и компьютерных данных

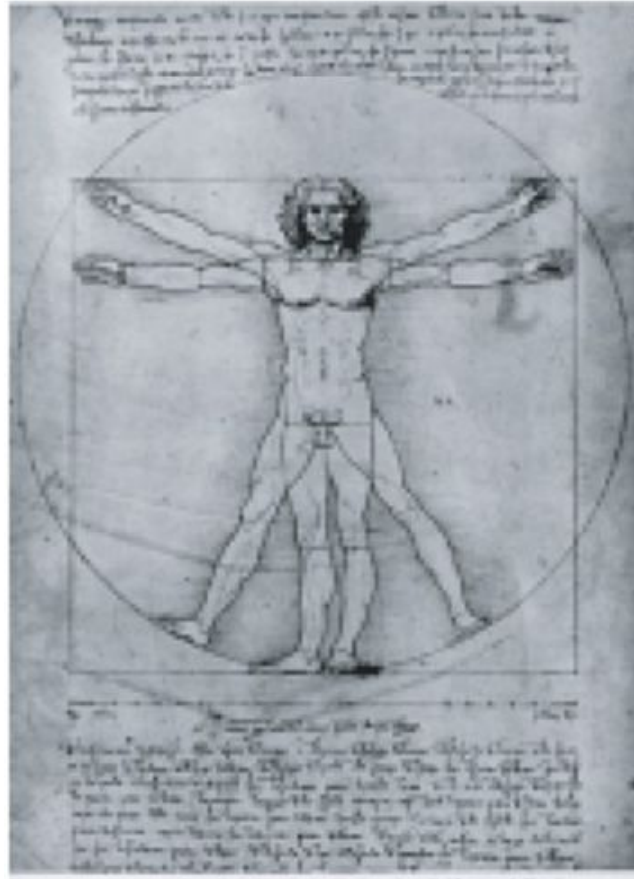
Base pairs	Unit	Abbreviation	Example
1	1 base pair	1 bp	
1000	1 kilobase pair	1 kb	Size of a typical coding region of a gene
1,000,000	1 megabase pair	1 Mb	Size of a typical bacterial genome
10^9	1 gigabase pair	1 Gb	The human genome is 3 billion base pairs
10^{12}	1 terabase pair	1 Tb	
10^{15}	1 petabase pair	1 Pb	

Size	Abbreviation	No. bytes	Examples
Bytes	–	1	1 byte is typically 8 bits, used to encode a single character of text
Kilobytes	1 kb	10^3	Size of a text file with up to 1000 characters
Megabytes	1 MB	10^6	Size of a text file with 1 million characters
Gigabytes	1 GB	10^9	600 GB: size of GenBank (uncompressed flat files) ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt (WebLink 2.84)
Terabytes	1 TB	10^{12}	385 TB: United States Library of Congress web archive (http://www.loc.gov/webarchiving/faq.html) (WebLink 2.85) 464 TB: Data generated by the 1000 Genomes Project (http://www.1000genomes.org/faq/how-much-disk-space-used-1000-genomes-project) (WebLink 2.86)
Petabytes	1 PB	10^{15}	1 PB: size of dataset available from The Cancer Genome Atlas (TCGA) 5 PB: size of SRA data available for download from NCBI 15 PB: amount of data produced each year at the physics facility CERN (near Geneva) (http://home.web.cern.ch/about/computing) (WebLink 2.87)
Exabytes	1 EB	10^{18}	2.5 exabytes of data are produced worldwide (Lampitt, 2014)

Увеличение количества известных последовательностей ДНК



Второй уровень: организм

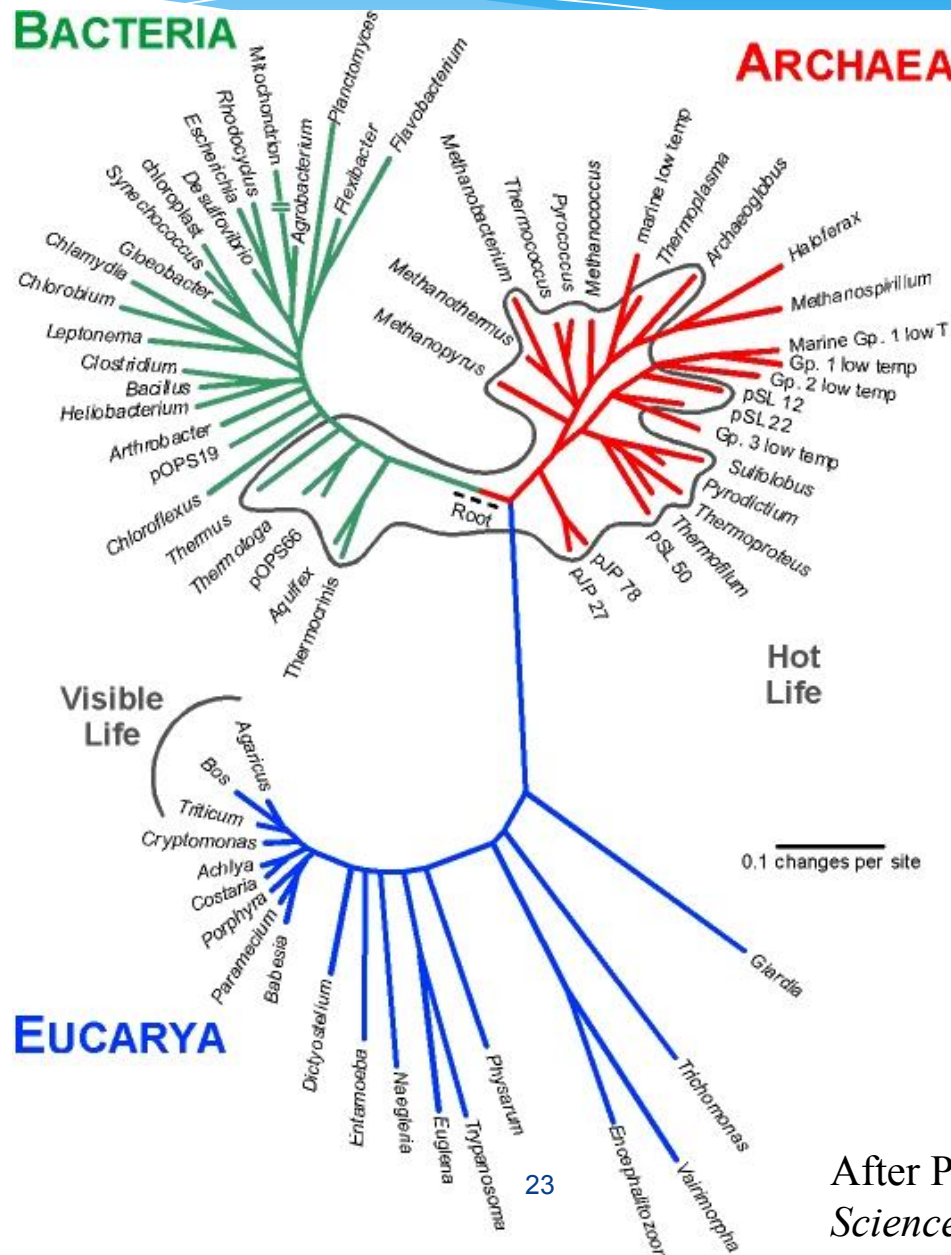


Время
развития

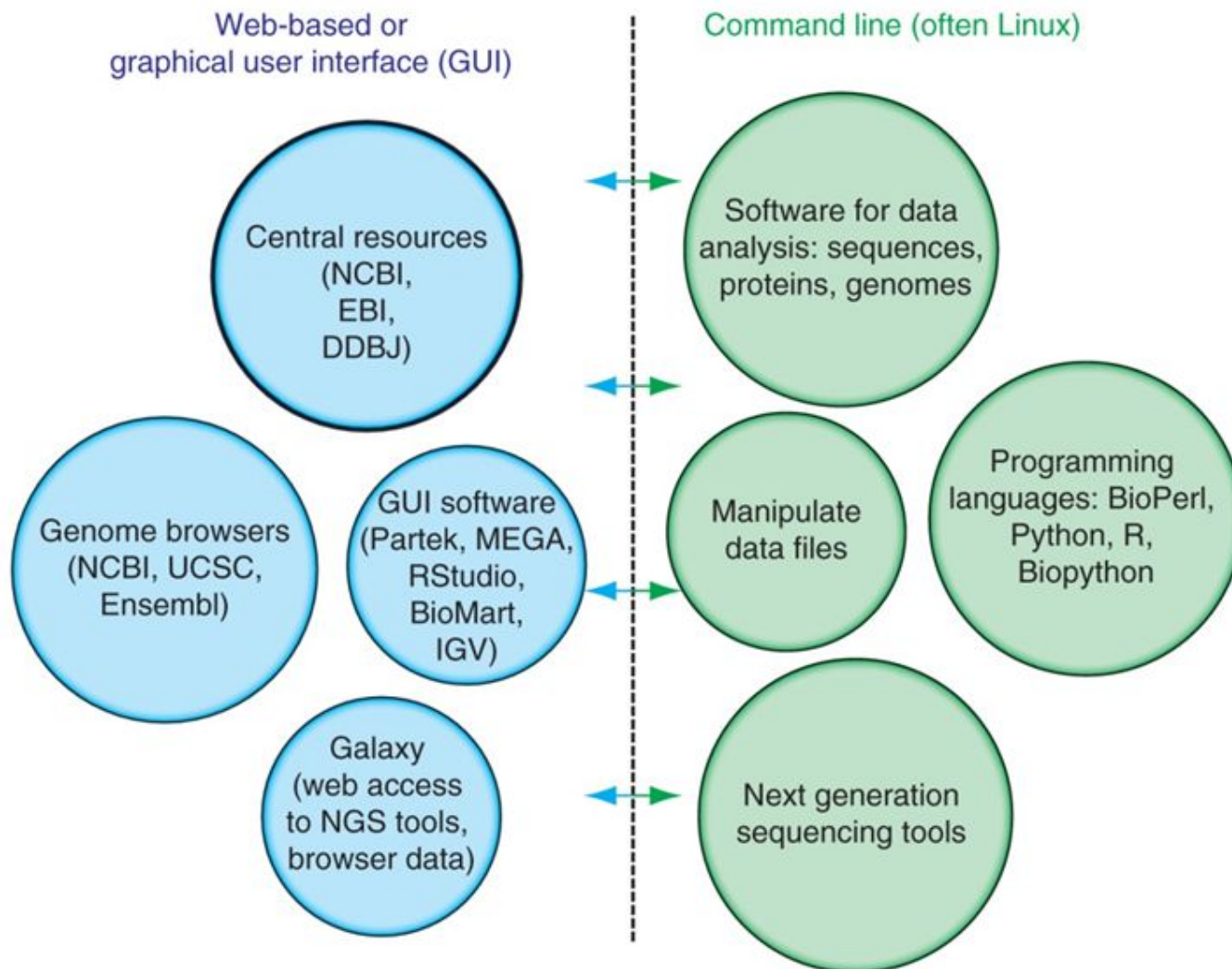


Ткани и органы, физиология,
фармакология, патология

Третий уровень: древо жизни



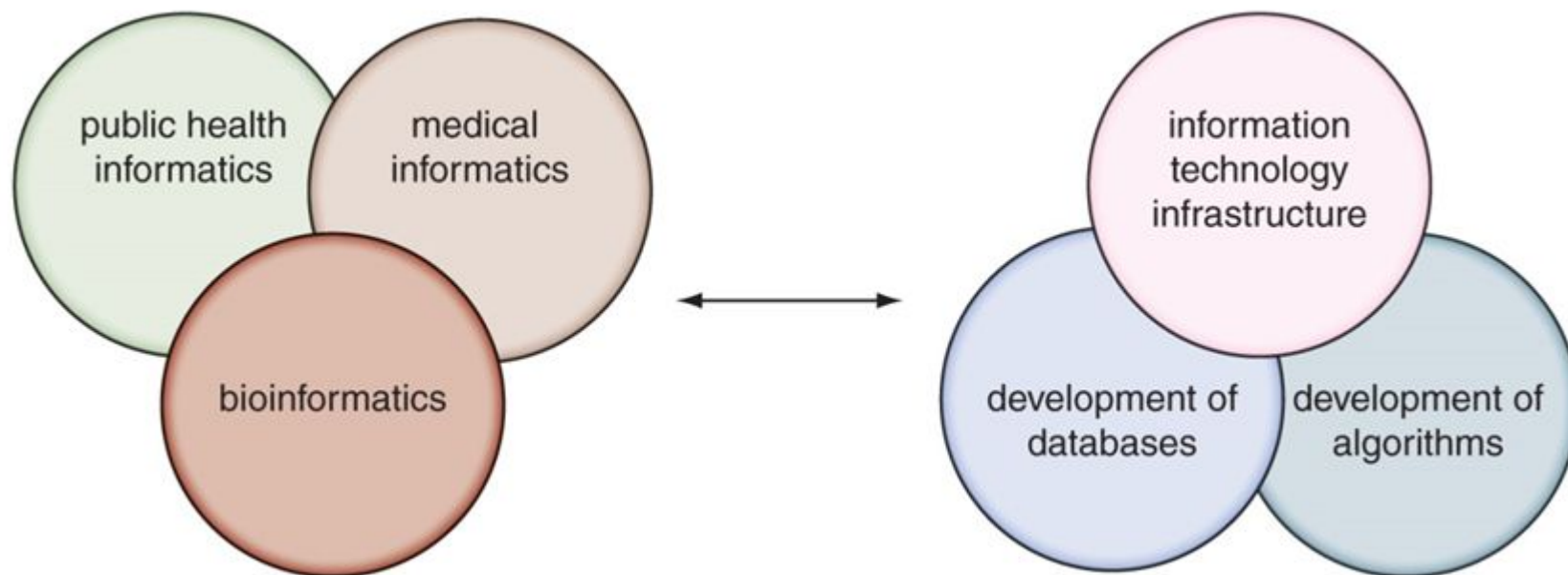
Ресурсы и инструменты



Компьютерные инструменты

пользователи

разработчики





Биологические базы данных

База данных — совокупность данных, систематизированных таким образом, чтобы они могли быть найдены и обработаны с помощью ЭВМ

Приложения и назначение биологических баз данных:

Задачи:

- * Проведение массивных вычислений, анализ и предсказание в любой области исследовательского интереса (Например, изучение последовательности белка, его структурный анализ, термостабильность, специфичность связывания его с ДНК и т.п.)
- * Проверка доступности данных для любых новых научных исследований
- * Сравнение экспериментальных данных с существующими данными
- * ...

Проблемы

- * Увеличивается потребность для использования баз, содержащих биологические данные -> необходима **обработка**
- * **Типы** биологических **данных** сильно **варьируют** (нуклеотиды, белки, геномы, таксономии и т.п.)
- * **Содержимое** в однотипных базах данных может **быть представлено по разному**, иметь разные форматы файлов и т.п. -> необходима адаптация данных
- * БД могут значительно отличаться по структуре – необходимы **интерфейсы**

Типы данных

- ✓ **Текст.** Примеры текстовых баз данных **PubMed** или **OMIM**, содержащие текстовую информацию и ссылки, связанные с биологическими данными.
- ✓ **Числовые данные.** Например, данные о генной экспрессии так же как другие данные получаемые с использованием microarray-технологии. Пример: база данных **ArrayExpress** европейского Института Биоинформатики (**EBI**) или **GEO** в **NCBI**.

Типы данных

- ✓ **Связи.** Большинство баз данных содержит информацию относительно каких-либо последовательностей в пределах определенной области интереса или предмета изучения. Различные типы баз данных это, например, **InterPro** база данных, состоящая из коллекции связей белковых доменов и белковых семейств с другими базами данных, представляющими собой связанные ресурсы.
- ✓ **Изображения.** В области изучения двумерного геля-электрофореза или микроскопических изображений Существуют различные базы данных, содержащие данные о, например, идентификации элемента геля на референтном или изучаемом изображении.

Типы данных

- ✓ **Данные о последовательности.** **GenBank** и **UniProt** – примеры биологических баз данных, содержащих ДНК и последовательности белка, соответственно
- ✓ **Структура белка.** Например, базы данных связанные с определенными файлами структуры белка как **PDB**, **SCOP** и **CATH** базы данных

Коллекции баз данных

- * Журнал *Nucleic acid research. Database issue* (первый выпуск каждого года) публикует список баз данных, покрывающих различные аспекты молекулярной и клеточной биологии.
 - * Он доступен для свободного доступа по адресу <http://nar.oupjournals.org/>
- + http://en.wikipedia.org/wiki/Biological_database

Коллекции баз данных

* <http://www.oxfordjournals.org/nar/database/a/>

The screenshot displays the Nucleic Acids Research (NAR) website interface. At the top, the 'OXFORD JOURNALS' logo is on the left, and navigation links for 'CONTACT US', 'MY BASKET', and 'MY ACCOUNT' are on the right. The main header features the journal title 'Nucleic Acids Research' in a large, white serif font against a dark red background. Below this, a horizontal navigation bar contains links: 'ABOUT THIS JOURNAL', 'CONTACT THIS JOURNAL', 'SUBSCRIPTIONS', 'CURRENT ISSUE', 'ARCHIVE', and 'SEARCH'. The main content area shows a breadcrumb trail: 'Oxford Journals > Life Sciences > Nucleic Acids Research > Database Summary Paper Alpha List'. The title '2014 NAR Database Summary Paper Alphabetical List' is prominently displayed. To the right of the title is a box with a list of links: 'Compilation Paper', 'Category List', 'Alphabetical List', 'Category/Paper List', and 'Search Summary Papers'. Below the title, an alphabetical index '1 2 3 4 5 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z' is provided. Three database entries are listed: 1. '1000 Genomes Selection Browser' by Engelken, Johannes; Pybus, Marc; Dall'Olio, Giovanni; Luisi, Pierre; Uzkudun, Manu; Carreño-Torres, Angel; Pavlidis, Pavlos; Laayouni, Hafid; Bertranpetit, Jaume. Signature of selection in the human genomes. Links: database, summary. 2. '16S and 23S Ribosomal RNA Mutation Database' by Triman K.L. 16S and 23S ribosomal RNA mutations. Links: database, summary. 3. '2D-PAGE' by Pleissner, K.-P., Eifert, T., Buettner, S., Knipper, J., Schmelzer, P., Stein, R., Schmidt, F., Mattow, J., Zimny-Arndt, U., Schmid, M., Jungblut, P.R. Proteome database system for microbial research. Links: database, summary.

Категории баз данных

- * Последовательности нуклеотидов
- * Последовательности РНК
- * Белковые последовательности
- * Структуры
- * Геномика (беспозвоночные)
- * Метаболические и сигнальные пути
- * Геномы человека и других позвоночных
- * Гены человека и заболевания
- * Данные экспериментов микромассивов (Microarray) и другие базы по экспрессии генов
- * Протеомные ресурсы
- * Иные молекулярно-биологические базы данных
- * Клеточные органеллы
- * Растения
- * Иммунология
- * Клеточная биология



<http://www.insdc.org/>



<http://www.ncbi.nlm.nih.gov/>



<http://www.ddbj.nig.ac.jp/>



<http://www.ebi.ac.uk/>



Обмен последовательностями между всеми базами данных нуклеотидных последовательностей производится ежедневно, поэтому в каждой из них постоянно содержится актуальная информация

National Center for Biotechnology Information (NCBI)

www.ncbi.nlm.nih.gov

NCBI

Resources

How To

NCBI

National Center for
Biotechnology Information

All Databases

Search

Sign in to NCBI

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation


Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News](#) | [Blog](#)


Submit

Deposit data or manuscripts into NCBI databases




Download

Transfer NCBI data to your computer




Learn

Find help documents, attend a class or watch a tutorial




Develop

Use NCBI APIs and code libraries to build applications




Analyze

Identify an NCBI tool for your data analysis task



Research

Explore NCBI research and collaborative projects



Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[PubMed Health](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI Announcements

Variation Viewer 1.5 adds facet toggling, updated backend data

04 Feb 2016

Variation Viewer 1.5 provides several new features, improvements and bug fixes

February 17th webinar: "Five ways to submit next-gen sequencing data to NCBI's Sequence Read Archive (SRA)"

03 Feb 2016

In two weeks, NCBI will present a webinar

Genome Workbench 2.10 now available

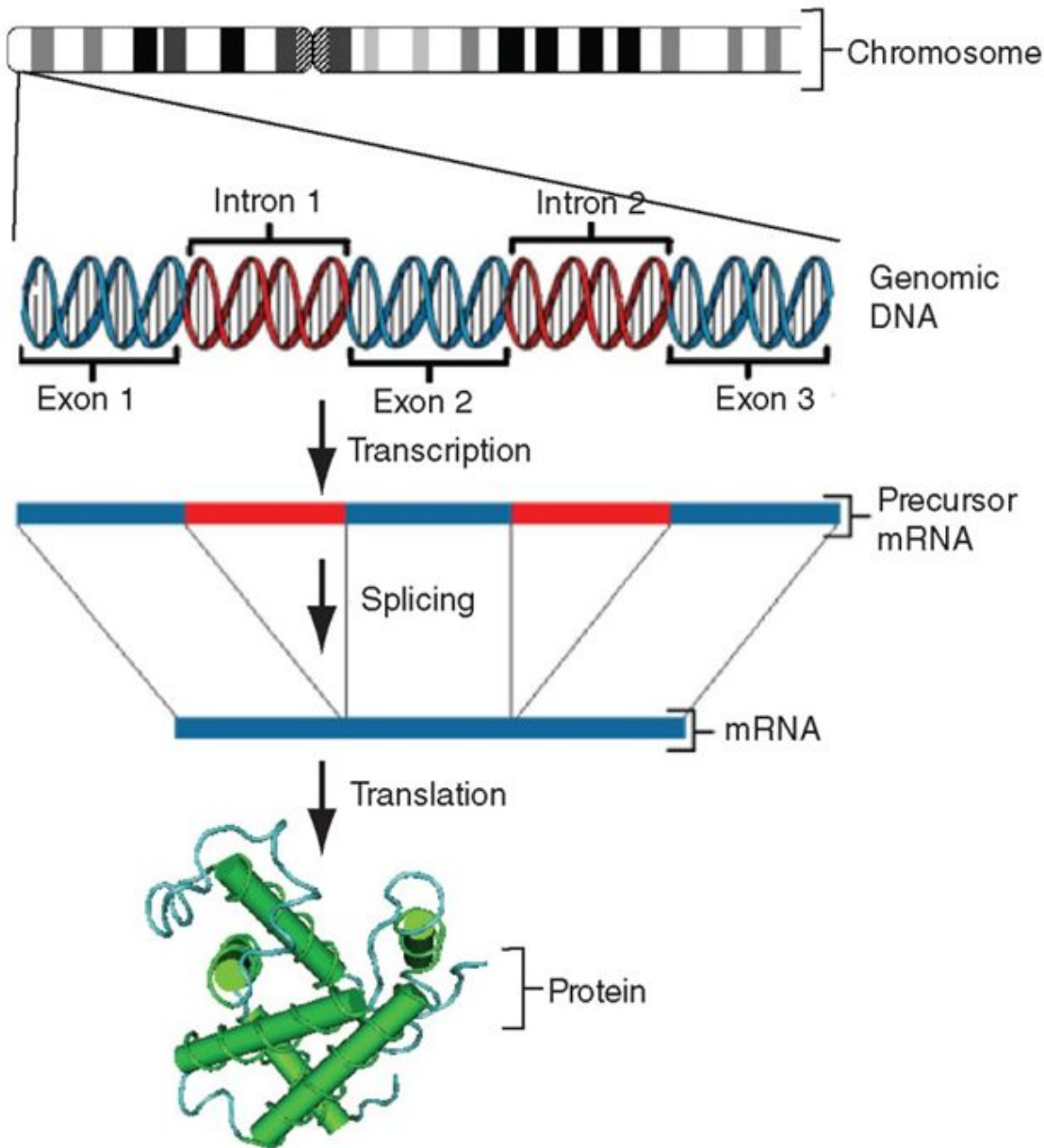
29 Jan 2016

Genome Workbench 2.10 includes a reworked BLAST tool and new functionalities in Tree View. For the full list

PHIMY

36

Какие данные?



Genome
dbVar

GenBank
SRA
dbGSS
dbHTGS
UniSTS
dbSNP

dbEST
UniGene
GEO profiles
GEO datasets

UniProt
Protein Data Bank
Conserved Domain Database

Ключевые особенности NCBI: PubMed

- Поисковый сервис National Library of Medicine (NLM)
- 28 млн. ссылок в MEDLINE (2018)
- связь с online журналами
- Как использовать PubMed:

<http://www.ncbi.nlm.nih.gov/pubmed>

или через NLM:

<http://www.nlm.nih.gov/bsd/disted/pubmed.html>

Ключевые особенности NCBI: поисковая система *gQuery*

***gQuery* глобальная система поиска во всех базах NCBI интегрирует...**

- научную литературу;
- базы данных по последовательностям ДНК и белков;
- данные по 3D структурам белков;
- наборы данных популяционных исследований;
- сборки полных геномов.

beta globin

Search

Results found in 34 databases for "beta globin"

Literature

Books	370	books and reports
MeSH	4	ontology used for PubMed indexing
NLM Catalog	11	books, journals and more in the NLM Collections
PubMed	9,594	scientific & medical abstracts/citations
PubMed Central	24,261	full-text journal articles

Health

ClinVar	114	human variations of clinical significance
dbGaP	1,921	genotype/phenotype interaction studies
GTR	20	genetic testing registry
MedGen	18	medical genetics literature and links
OMIM	121	online mendelian inheritance in man
PubMed Health	30	clinical effectiveness, disease and drug reports

Genomes

Assembly	0	genome assembly information
BioProject	37	biological projects providing data to NCBI
BioSample	21	descriptions of biological source materials
Clone	32,086	genomic and cDNA clones
dbVar	75	genome structural variation studies
Genome	145	genome sequencing projects by organism
GSS	3	genome survey sequences
Nucleotide	4,613	DNA and RNA sequences
Probe	125	sequence-based probes and primers
SNP	995	short genetic variations
SRA	104	high-throughput DNA and RNA sequence read archive
Taxonomy	0	taxonomic classification and nomenclature catalog

Genes

EST	2,042	expressed sequence tag sequences
Gene	138	collected information about gene loci
GEO Data Sets	171	functional genomics studies
GEO Profiles	5,246	gene expression and molecular abundance profiles
HomoloGene	4	homologous gene sets for selected organisms
PopSet	79	sequence sets from phylogenetic and population studies
UniGene	41	clusters of expressed transcripts

Proteins

Conserved Domains	9	conserved protein domains
Protein	3,026	protein sequences
Protein Clusters	0	sequence similarity-based protein clusters
Structure	1,153	experimentally-determined biomolecular structures

Chemicals

BioSystems	342	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	42	bioactivity screening studies
PubChem Compound	0	chemical information with structures, information and links
PubChem Substance	249	deposited substance and chemical information

Номер доступа (accession number) это обозначение для последовательности

NCBI включает базы данных (такие как GenBank), которые содержат информацию о последовательностях ДНК, РНК и белков

Вы можете получить информацию, начиная с такого запроса как имя интересующего белка или гена, последовательность нуклеотидов ДНК, представляющих интерес

Последовательности ДНК и другие молекулярные данные помечены **номерами доступа**, которые используются для идентификации последовательности или иной записи, относящейся к данным о молекулах

Что такое номер доступа?

Номер доступа (*accession number*) это метка, используемая для идентификации последовательности.

Это строка из букв и/или цифр, которая однозначно соответствует молекулярной последовательности.

Примеры:

X02775 GenBank genomic DNA sequence

NG_000007.3 RefSeqGene

rs192792910 dbSNP (single nucleotide polymorphism)

DNA

AA970968.1 An expressed sequence tag (EST) (1 of 2,345)

NM_000518.4 RefSeq DNA sequence (from a transcript)

RNA

NP_000509.1 RefSeq protein

CAA00182.1 GenBank protein

Q14473 SwissProt protein

1YEO|B Protein Data Bank structure record

protein

RefSeq проект NCBI:

«лучшие представители» последовательностей

RefSeq (доступен через главную страницу NCBI)
Обеспечивает доступ к экспертно курированной последовательности, которая соответствует ее самой *стабильной, согласованной «эталонной»* версии.

RefSeq идентификаторы включают следующие форматы:

Complete genome NC_#####

Complete chromosome NC_#####

Genomic contig NT_#####

mRNA (DNA format) NM_##### e.g. NM_000518

Protein NP_##### e.g. NP_000509

Доступ к последовательности:

Entrez Gene NCBI

Entrez Gene содержит ключевую информацию по каждому гену/белку из основных баз данных NCBI.

RefSeq предоставляет *курируемый* номер доступа для каждой ДНК (NM_000518.4 для ДНК бета глобина, соответствующей мРНК) или белка (NP_000509.1).

Введем “beta globin” На главной
странице NCBI и нажмем “Search”

NCBI Resources ▾ How To ▾ Sign in to NCBI

NCBI
National Center for
Biotechnology Information

All Databases ▾ beta globin × Search

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Visualization

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS Feeds](#)

Get Started

- [Tools](#): Analyze data using NCBI software
- [Downloads](#): Get NCBI data or software
- [How-To's](#): Learn how to accomplish specific tasks at NCBI
- [Submissions](#): Submit data to GenBank or other NCBI databases

Popular Resources

PubMed

Bookshelf

PubMed Central

PubMed Health

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

NCBI Facebook page

Find out the latest news about NCBI resources and participate in community discussions.

GO



NCBI Announcements

New version of Genome Workbench available

08 Sep 2012

An integrated, downloadable application

Search NCBI databases

Help

About 79,887 search results for "beta globin"

Literature

Books	384	books and reports
MeSH	4	ontology used for PubMed indexing
NLM Catalog	10	books, journals and more in the NLM Collections
PubMed	9,044	scientific & medical abstracts/citations
PubMed Central	20,279	full-text journal articles

Health

ClinVar	183	human variations of clinical significance
dbGaP	1,577	genotype/phenotype interaction studies
GTR	20	genetic testing registry
MedGen	14	medical genetics literature and links
OMIM	119	online mendelian inheritance in man
PubMed Health	21	clinical effectiveness, disease and drug reports

Genomes

Assembly	0	genomic assembly information
BioProject	27	biological projects providing data to NCBI
BioSample	21	descriptions of biological source materials
Clone	32,086	genomic and cDNA clones
dbVar	245	genome structural variation studies
Epigenomics	24	epigenomic studies and display tools
Genome	389	genome sequencing projects by organism
GSS	3	genome survey sequences

Genes

EST	2,042	expressed sequence tag sequences
Gene	118	collected information about gene loci
GEO DataSets	153	functional genomics studies
GEO Profiles	4,463	gene expression and molecular abundance profiles
HomoloGene	4	homologous gene sets for selected organisms
PopSet	64	sequence sets from phylogenetic and population studies
UniGene	41	clusters of expressed transcripts

Proteins

Conserved Domains	8	conserved protein domains
Protein	2,390	protein sequences
Protein Clusters	0	sequence similarity-based protein clusters
Structure	1,092	experimentally-determined biomolecular structures

Chemicals

BioSystems	319	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	46	bioactivity screening studies
PubChem Compound	0	chemical information with structures, information and links
PubChem Substance	211	deposited substance and chemical information

Идем по ссылке "Gene"

Результаты поиска в Gene

[Resources](#) [How To](#) [Sign in to NCBI](#)

Gene [Save search](#) [Advanced](#) [Help](#)

[Show additional filters](#) **Display Settings:** ☒ Tabular, 20 per page, Sorted by Relevance **Send to:** ☒ [Hide sidebar >>](#)

[Clear all](#) **Results: 1 to 20 of 110** [Filters activated: Current only. \[Clear all\]\(#\) to show 118 items.](#)

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> HBB ID: 3043	hemoglobin, beta	Chromosome 11, NC_000011.10 (5225466..5227071, complement)	CD113t-C, beta-globin	141900
<input type="checkbox"/> hbg1 ID: 394453	hemoglobin, gamma A	NW_004668244.1 (60116737..60118249)	beta-globin, hbb1, hbga, hbgr, hsggl1	
<input type="checkbox"/> hbg1 ID: 734881	hemoglobin, gamma A		beta-globin, hbb1, hbga, hbgr, hsggl1	
<input type="checkbox"/> Hbb-bh1 ID: 15132	hemoglobin Z, beta-like embryonic chain	Chromosome 7, NC_000073.6 (103841638..103843162, complement)	betaH1	
<input type="checkbox"/> HBG2 ID: 396485	hemoglobin, gamma G	Chromosome 1, NC_006088.3 (193724299..193725801)	HBB, HBD, HBE1	
<input type="checkbox"/> hbbe1.1 ID: 81538	hemoglobin beta embryonic-1.1	Chromosome 3, NC_007114.6 (54859510..54860272)	Sl:dZ118J2.7, Sl:dZ118J2.9, bE1, hbbe1, si:by119c3.2	
<input type="checkbox"/> HBB ID: 280813	hemoglobin, beta	Chromosome 15, AC_000172.1 (49022978..49024620)		
<input type="checkbox"/> hbb1 ID: 100136584	beta-globin			
<input type="checkbox"/> LOC689064 ID: 689064	beta-globin	Chromosome 1, NC_005100.4 (168964202..168965568)		
<input type="checkbox"/> HBB	hemoglobin, beta	Chromosome 11, NC_006478.3		

Gene sources
Genomic

Categories
Alternatively spliced
Annotated genes
Non-coding
Protein-coding
Pseudogene

Sequence content
CCDS
Ensembl
RefSeq
RefSeqGene

Status
☒ **Current only**

[Chromosome locations](#)
Select ...

[Clear all](#) [Show additional filters](#)

Filters: [Manage Filters](#)
Top Organisms [\[Tree\]](#)
Homo sapiens (39)
Mus musculus (27)
Rattus norvegicus (5)
Gallus gallus (4)
Danio rerio (4)
All other taxa (31)
[More...](#)

Find related data
Database:
[Find items](#)

Search details
beta globin[All Fields] AND alive[property]
[Search](#) [See more...](#)

Recent activity
[Turn Off](#) [Clear](#)
Q beta globin AND (alive[property]) (110)
Gene

Entrez Gene:

Полезное резюме и ссылки на другие базы данных

Resources ▾ How To ▾ Sign in to NCBI

Gene [Advanced](#) [Help](#)

[Display Settings:](#) ▾ Full Report [Send to:](#) ▾ [Hide sidebar >>](#)

HBB hemoglobin, beta [*Homo sapiens* (human)]
Gene ID: 3043, updated on 1-Feb-2015

Summary

Official Symbol HBB provided by [HGNC](#)

Official Full Name hemoglobin, beta provided by [HGNC](#)

Primary source [HGNC:HGNC:4827](#)

See related [Ensembl:ENSG00000244734](#); [HPRD:00786](#); [MIM:141900](#); [Vega:OTTHUMG00000066678](#)

Gene type protein coding

RefSeq status REVIEWED

Organism [Homo sapiens](#)

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

Also known as CD113t-C; beta-globin

Summary The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia. The order of the genes in the beta-globin cluster is 5'-epsilon -- gamma-G -- gamma-A -- delta -- beta--3'. [provided by RefSeq, Jul 2008]

Orthologs [mouse](#) [all](#)

Genomic context

Location: 11p15.5 [See HBB in \[MapViewer\]\(#\)](#)

Exon count: 3

Annotation release	Status	Assembly	Chr	Location
PHIMU				

Table of contents

[Summary](#)

[Genomic context](#)

[Genomic regions, transcripts, and products](#)

[Bibliography](#)

[Phenotypes](#)

[Variation](#)

[Pathways from BioSystems](#)

[Interactions](#)

[General gene information](#)
[Markers, Related pseudogene\(s\), Homology, Gene Ontology](#)

[General protein information](#)

[NCBI Reference Sequences \(RefSeq\)](#)

[Related sequences](#)

[Additional links](#)
[Locus-specific Databases](#)

Related information

[Order cDNA clone](#)

[3D structures](#)

[BioAssay](#)

[BioAssay by Target \(List\)](#)

[BioAssay by Target \(Summary\)](#)

[BioAssay, by Gene target](#)

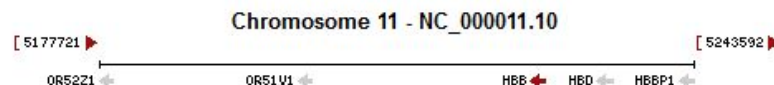
Genomic context

Location: 11p15.5

See HBB in [Epigenomics](#), [MapViewer](#)

Exon count: 3

Annotation release	Status	Assembly	Chr	Location
107	current	GRCh38.p2 (GCF_000001405.28)	11	NC_000011.10 (5225466..5227071, complement)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	11	NC_000011.9 (5246696..5248301, complement)

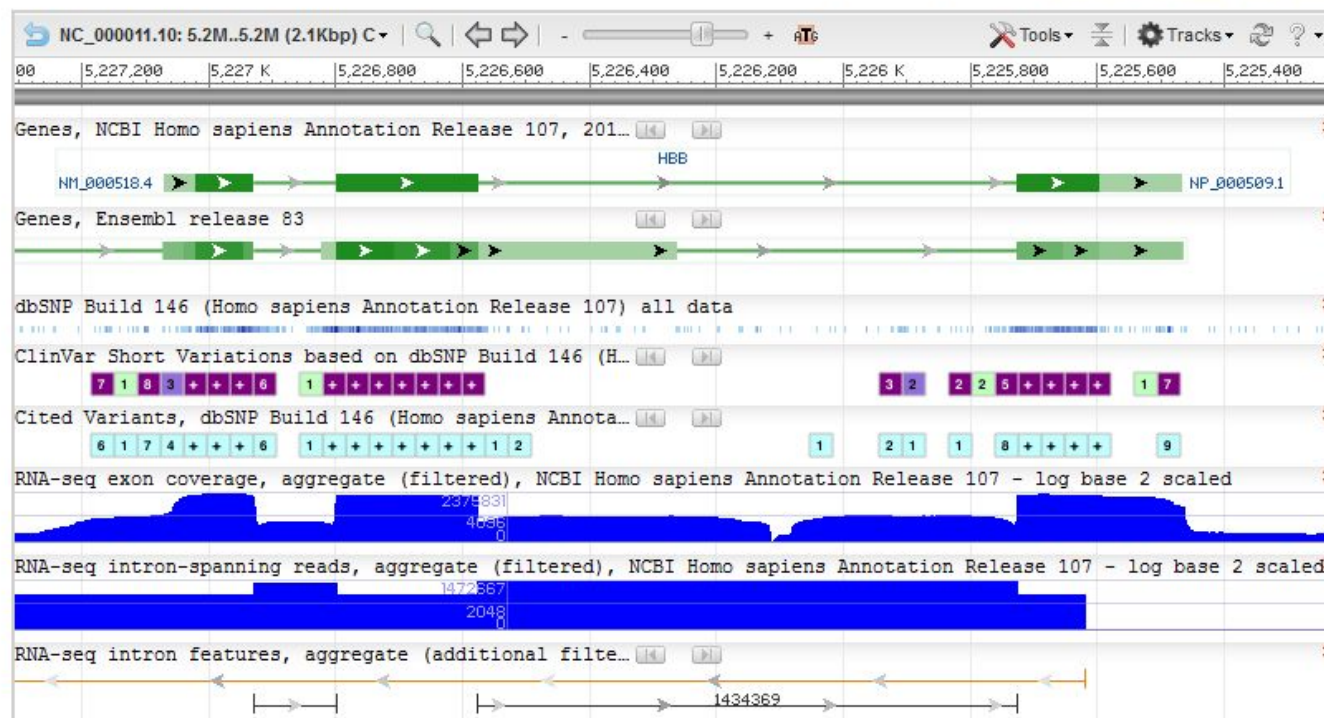


Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence: NC_000011.10 Chromosome 11 Reference GRCh38.p2 Primary Assembly

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)



Order cDNA clone

3D structures

BioAssay

BioAssay by Target (List)

BioAssay by Target (Summary)

BioAssay, by Gene target

BioAssays, RNAi Target, Tested

BioProjects

BioSystems

Books

CCDS

ClinVar

Conserved Domains

dbVar

Full text in PMC

Full text in PMC_nucleotide

Gene neighbors

Genome

GEO Profiles

GTR

HomoloGene

Map Viewer

MedGen

Nucleotide

OMIM

Probe

Protein

PubChem Compound

PubChem Substance

PubMed

PubMed (GeneRIF)

PubMed (OMIM)

PubMed(nucleotide/PMC)

RefSeq Proteins

RefSeq RNAs

RefSeqGene

Страница “Gene” NCBI предоставляет следующую полезную информацию

- Геномный контекст
- Библиография
- Фенотипическое проявление
- Gene Ontology (принципы организации биологических процессов, молекулярная функция, клеточный компонент)
- Референтные последовательности
- Дополнительные (не RefSeq последовательности)
- Много, много ссылок на NCBI ресурсы (e.g. HomoloGene)
- Много ссылок на внешние ресурсы

Entrez Protein:
accession,
organism,
literature...

NCBI Resources How To

Protein

Protein

Advanced

Display Settings: GenPept

Send to:

Change region shown

Customize view

hemoglobin, beta [Homo sapiens]

GenBank: ADL14493.1

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

Go to:

LOCUS ADL14493 147 aa linear PRI 10-AUG-2010

DEFINITION hemoglobin, beta [Homo sapiens].

ACCESSION ADL14493

VERSION ADL14493.1 GI:302313137

DBSOURCE accession [GU324922.1](#)

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)

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

REFERENCE 1 (residues 1 to 147)

AUTHORS Rieder,M.J., Bertucci,C., Stanaway,I.B., Johnson,E.J.,
Swanson,J.E., Siegel,D.L., da Ponte,S.H., Igartua,C., Patterson,K.
and Nickerson,D.A.

TITLE Direct Submission

JOURNAL Submitted (25-NOV-2009) Genome Sciences, University of Washington,
1705 NE Pacific, Seattle, WA 98195, USA

COMMENT Method: conceptual translation supplied by author.

FEATURES Location/Qualifiers

source 1..147
/organism="Homo sapiens"
/db_xref="taxon:9606"

Protein 1..147
/product="hemoglobin, beta"

Region 7..146
/region_name="Hb-beta_like"
/note="Hemoglobin beta, gamma, delta, epsilon, and related
Hb subunits; cd08925"

GI (GenInfo Identifier)

Analyze this sequence

[Run BLAST](#)

[Identify Conserved Domains](#)

[Highlight Sequence Features](#)

[Find in this Sequence](#)

Protein 3D Structure



Structure Of T. Brucei
Haptoglobin-hemoglobin
Receptor Binding To
PDB: 4WJG
Source: Homo sapiens,
Staphylococcus aureus
subsp. aureus N315,

Trypanosoma brucei brucei

Method: X-Ray Diffraction

Resolution: 3.1 Å

[See all 210 structures...](#)

Articles about the HBB gene

Induction of fetal hemoglobin through enhanced
translation efficiency of γ -globin m [Blood. 2014]

β -Globin gene sequencing of hemoglobin
Austin revises the I [Arch Pathol Lab Med. 2014]

Case report: prenatal diagnosis of Hb
Hammersmith [B42(CD1)PI [Hemoglobin. 2014]

Entrez Protein:

...особенности белка и его последовательность в однобуквенном коде

Site

```
94
/site_type="modified"
/experiment="experimental evidence, no additional details
recorded"
/note="S-nitrosylation site"
/citation=[7]
```

Site

```
121
/site_type="glycosylation"
/experiment="experimental evidence, no additional details
recorded"
/note="glycation site"
/citation=[9]
```

CDS

```
1..147
/gene="HBB"
/gene_synonym="CD113t-C"
/coded_by="NM_000518.4:51..494"
/db_xref="CCDS:CCDS7753.1"
/db_xref="GeneID:3043"
/db_xref="HGNC:4827"
/db_xref="HPRD:00786"
/db_xref="MIM:141900"
```

ORIGIN

```
1 mvhltpeeks avtalwgkvn vdevggealg rllvvypwtq rffesfgdls tpdavmgnpk
61 vkahgkkvlg afsdglahld nlkgtfatls elhcdklhvd penfrllgnv lvcvlahhfg
121 keftppvqaa yqkvvagvan alahkyh
```

//

Название аминокислот и их трех- и однобуквенный код

Name	3-Letter	1-Letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I

Name	3-Letter	1-Letter
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Entrez Protein:

Можно изменить вид предоставляемой информации

NCBI Resources ☒ How To ☒ My NCBI Sign In

Protein
Translations of Life

Search: Protein Limits Advanced search Help

Search Clear

Display Settings: ☒ GenPept

Send to: ☒

Change region shown

Customize view

hemoglobin subunit beta [Homo sapiens]

NCBI Reference Sequence: NP_000509.1

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

[Go to:](#) ☒

LOCUS NP_000509 147 aa linear PRI 05-NOV-2010

DEFINITION hemoglobin subunit beta [Homo sapiens].

ACCESSION NP_000509

VERSION NP_000509.1 GI:4504349

DBSOURCE REFSEQ: accession [NM_000518.4](#)

KEYWORDS .

SOURCE Homo sapiens (human)

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

REFERENCE 1 (residues 1 to 147)

AUTHORS Bailey,S.D., Xie,C., Do,R., Montpetit,A., Diaz,R., Mohan,V.,
Keavney,B., Yusuf,S., Gerstein,H.C., Engert,J.C. and Anand,S.

CONSRMT DREAM investigators

TITLE Variation at the NFATC2 locus increases the risk of
thiazolidinedione-induced edema in the Diabetes REDuction

Analyze this sequence

Run BLAST

Identify Conserved Domains

Find in this Sequence

Articles about the HBB gene

Transfusion independence and HMGA2
activation after gene therapy of hui [Nature. 2010]

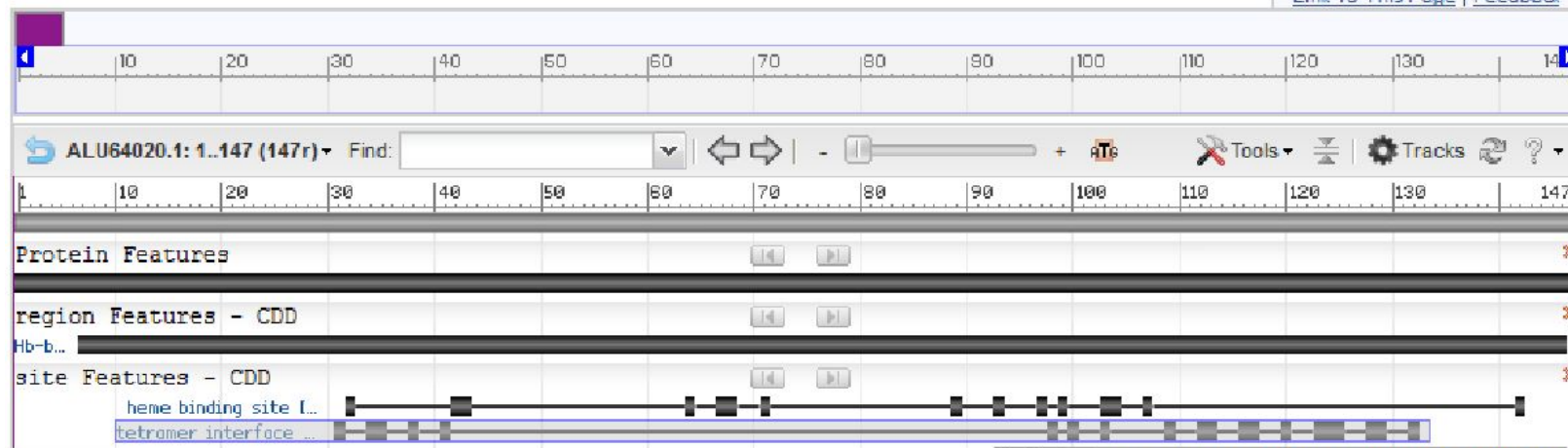
Secondary erythrocytosis caused by hemoglobin
Tak(δβ)0-thalassemia sy [Acta Haematol. 2010]

Haemoglobin O in An Arab Family.
[Br Med J. 1960]

See all...

beta-globin [Homo sapiens]

GenBank: ALU64020.1

[GenPept](#) [Identical Proteins](#) [FASTA](#)[Link To This Page](#) | [Feedback](#)

tetramer interface [polypeptide binding]
site: other
Title: tetramer interface [polypeptide binding]
Location: 31..132
[Length]
Span: 102
Placed: 20

Links & Tools
View CDD: [271276](#)

GenBank View: [ALU64020.1 \(31..132\)](#)
FASTA View: [ALU64020.1 \(31..132\)](#)
BLAST protein: [ALU64020.1 \(31..132\)](#)

FASTA формат:

NCBI Resources How To My NCBI

Protein
Translations of Life

Search: Protein Limits Advanced search Help

Search Clear

Display Settings: FASTA

Send to:

Change region shown

hemoglobin subunit beta [Homo sapiens]

NCBI Reference Sequence: NP_000509.1

[GenPept](#) [Graphics](#)

```
>gi|4504349|ref|NP_000509.1| hemoglobin subunit beta [Homo sapiens]
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVGMGNPKVKAHGKKVLG
AFSDGLAHLDDLKGTFAITLSELHCDKLHVDPENFRLLGNVLCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH
```

Analyze this sequence

[Run BLAST](#)

[Identify Conserved Domains](#)

[Find in this Sequence](#)

Articles about the HBB gene

**Универсальный, компактный текстовый формат.
После одной строки заголовка следуют строки нуклеотидов
или аминокислот в однобуквенном коде**

Форматы представления данных

FASTA	однобуквенная последовательность ДНК или белка
FASTQ	ДНК последовательность с оценкой качества секвенирования каждого основания
SAM	Sequence Alignment/Map file (tab-delimited)
BAM	сжатый двоичный вариант SAM
VCF	variant call format (genomic variants; indels)

(См. genome.ucsc.edu/FAQ/FAQformat.html для следующих форматов:)

BED	(Browser Extensible Data)a table including chromosome, start, end
WIG	wiggle format (displays dense, continuous data)
GFF	General Feature Format (tab separated)

Excel (.xls, .xlsx) подобные табличные форматы:

.txt	tab-delimited text file (or space delimited)
.csv	comma separated text file

FASTQ формат

Состоит из 4-х линий

```
@EAS54_6_R1_2_1_413_324
CCCTTCTTGTCTTCAGCGTTTCGCC
+
::3::::::::::::7:::::::::88
@EAS54_6_R1_2_1_540_792
TTGGCAGGCCAAGGCCGATGGATCA
+
::::::::::::7::::::-:::3;83
@EAS54_6_R1_2_1_443_348
GTTGCTTCTGGCGTGGGTGGGGGGG
+EAS54_6_R1_2_1_443_348
::::::::::::9;7:::7;393333
```

Идентификатор последовательности
последовательность (как FASTA)

Оценка качества (per base)

Genome Browsers

- * Геномная ДНК организована в хромосомы. Геномные браузеры отображают идеограммы (картинки) хромосом, с выбранным пользователем «треками аннотаций», которые отражают различные виды информации (применительно к локализации области интереса в хромосоме).
- * Три наиболее важных геномных браузера:
 - * NCBI Map Viewer,
 - * Ensembl
 - * UCSC (University of California, Santa Cruz).



Human genome overview page (Annotation Release 106)

Human genome overview page (Annotation Release 105)

[Map Viewer Home](#)

Map Viewer Help
Human Maps Help
FTP

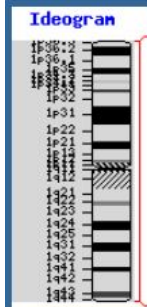
Data As Table View

[Maps & Options](#)

Region Shown:



You are here:



☒ default
☐ master

[PubMed](#)
[Entrez](#)
[BLAST](#)
[OMIM](#)
[Taxonomy](#)
[Structure](#)

[Search](#)

[Homo sapiens \(human\)](#) [Annotation Release 106 \(Current\)](#)

[BLAST human sequences](#)

Chromosome: [1] 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT

Master Map: Genes On Sequence

[Summary of Maps](#)

[Maps & Options](#)

Region Displayed: 0-250M bp

[Download/View Sequence/Evidence](#)

Ideogram	Hs UniG	Genes_seq	Symbol	O	Links	E	Cyto	Description
	Hs.735900 Hs.397705		TP73	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p36.3	tumor protein p73
			MTOR	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p36.2	mechanistic target of rapamycin (
			MTHFR	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p36.3	OTTHUMP00000002368
			NPPB	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p36.2	natriuretic peptide B
	Hs.679809 Hs.477370		HDAC1	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p34	histone deacetylase 1
	Hs.146542 Hs.740425		JUN	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p32-p31	jun proto-oncogene
			LEPR	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p31	leptin receptor
	Hs.732392 Hs.732666 Hs.535034		F3	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p22-p21	coagulation factor III (thrombopl
	Hs.732334 Hs.732334		GSTM1	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p13.3	glutathione S-transferase mu 1
	Hs.732540 Hs.622632 Hs.622632		PTPN22	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1p13.2	protein tyrosine phosphatase, nor
	Hs.656782 Hs.571920 Hs.655246		MUC1	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q21	mucin 1, cell surface associated
	Hs.732540 Hs.515947 Hs.734095		LMNA	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q22	lamin A/C
	Hs.732431 Hs.740600 Hs.734876		CRP	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q23.2	C-reactive protein, pentraxin-rela
	Hs.585378 Hs.731785		F5	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q23	coagulation factor V (proacceleri
			FASLG	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q23	Fas ligand (TNF superfamily, me
	Hs.732307 Hs.740387 Hs.737516		PTGS2	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q25.2-q25.3	prostaglandin-endoperoxide syntl
	Hs.655974 Hs.740425		CFH	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q32	complement factor H
			IL10	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q31-q32	interleukin 10
	Hs.732315 Hs.109514		PARP1	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q41-q42	poly (ADP-ribose) polymerase 1
	Hs.736030 Hs.567571		AGT	+	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	1q42.2	angiotensinogen (serpin peptidas

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail**
 - Comparative Genomics
 - Alignments (image) (51)
 - Alignments (text) (51)
 - Multi-species view (47)
 - Synteny (12)
 - Genetic Variation
 - Resequencing (2)
 - Linkage Data
 - Markers
 - Other genome browsers
 - UCSC
 - NCBI

- Configure this page
- Manage your data
- Export data
- Bookmark this page

Вывод
Ensembl для
beta globin
включает вид
хромосомы
11 (вверху),
участок (в
середине), и
детальный
вид (внизу).

Различные горизонтальные
аннотационные треки

Chromosome 11: 5,246,694-5,250,625

chromosome 11

p15.4

p13

p12

q14.1

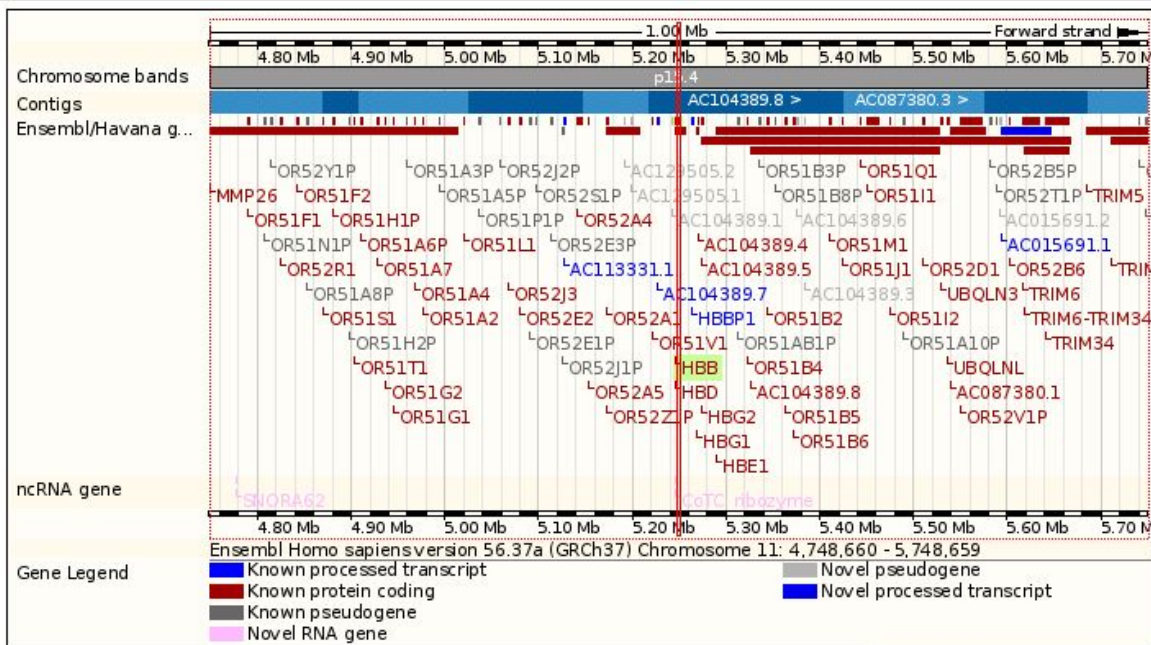
q22.3

Export Image

« Region overview

Region in detail [help](#)

Alignments (image) »

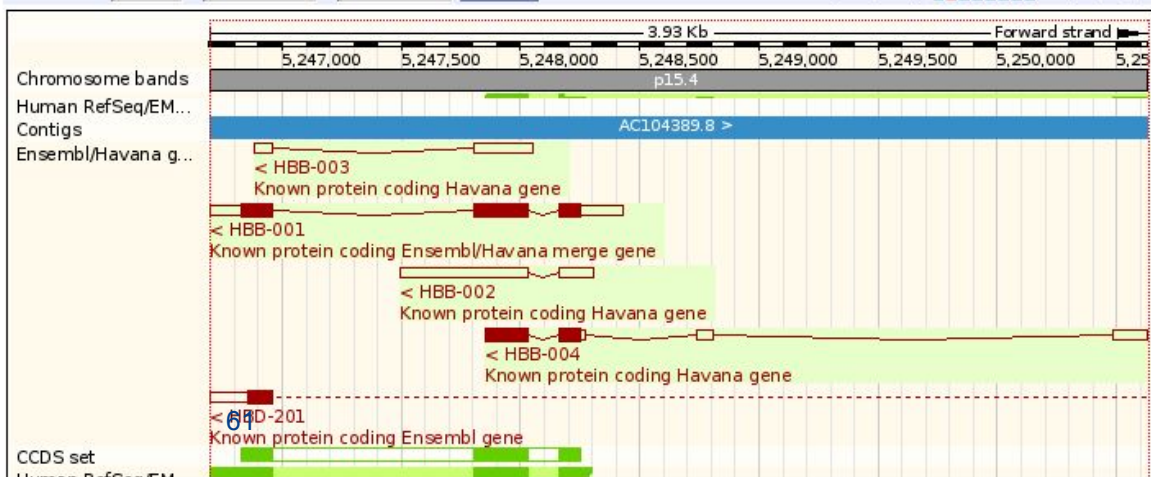


Export Image

Location: 11 : 5246694 - 5250625

Go>

« < > »



UCSC Genome Browser

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

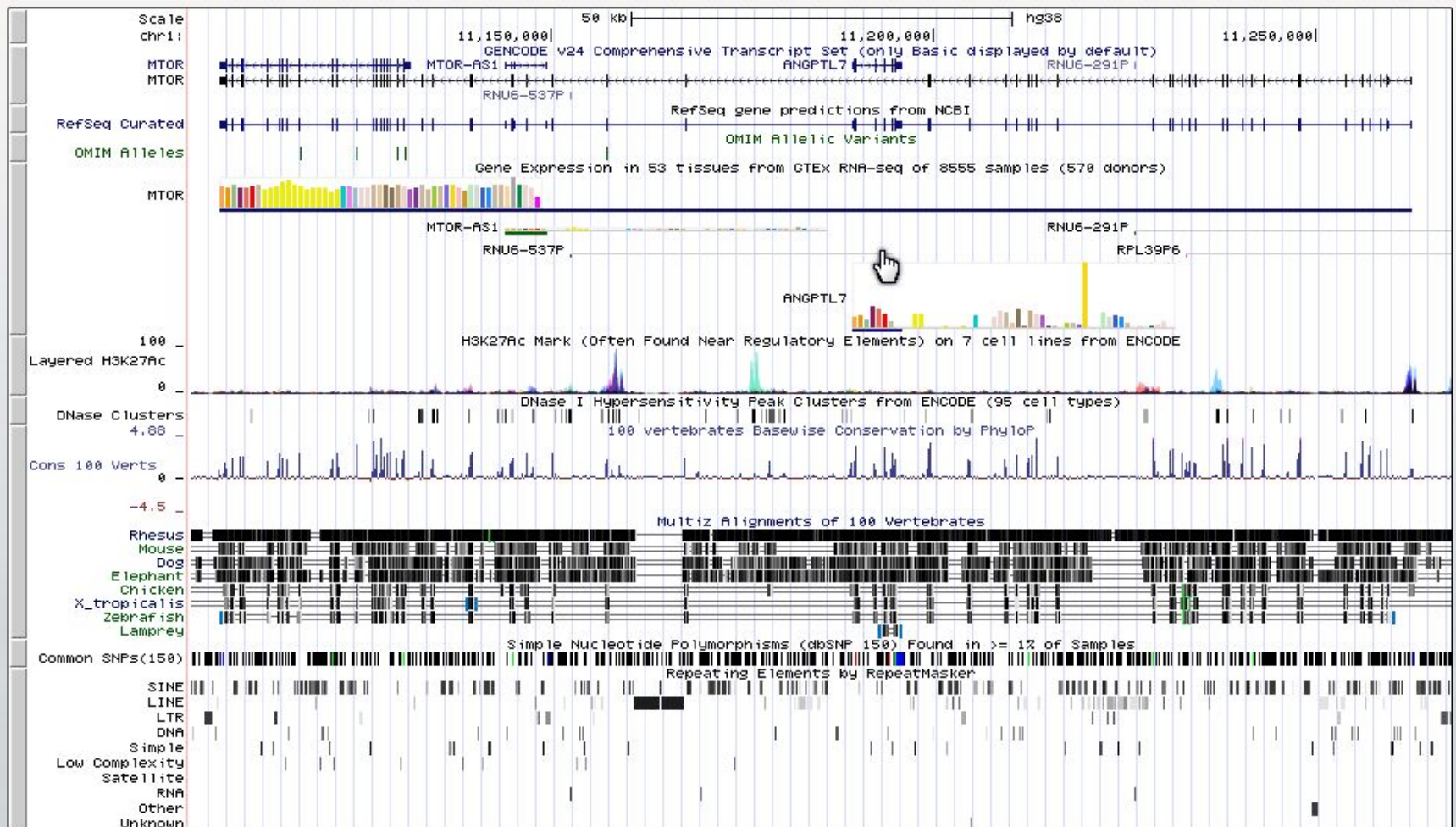
chr1:11,102,837-11,267,747 164,911 bp.

enter position, gene symbol, HGVS or search terms

go

[Request onsite workshops](#)

chr1 (p36.22) 1p31.1 1q12 1q41 1q44



The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB

UniProt Knowledgebase

Swiss-Prot (553,474)

Manually annotated and reviewed.

Records with information extracted from literature and curator-evaluated computational analysis.

TrEMBL (73,711,881)

Automatically annotated and not reviewed.

Records that await full manual annotation.

UniRef

The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records.

UniParc

UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.

Proteomes

A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.

Supporting data

Literature citations	Taxonomy	Subcellular locations
Cross-ref. databases	Diseases	Keywords

News

Forthcoming changes
Planned changes for UniProt

UniProt release 2017_01
Sheep in wolves' clothing | Change of the UniRef FASTA header

UniProt release 2016_11
From mouth to gut, a new mechanism for fimbriae assembly | Change of URIs for Ensembl and Ensembl Genomes

[News archive](#)

Getting started

- Text search**
Our basic text search allows you to search all the resources available
- BLAST**
Find regions of similarity between your sequences
- Sequence alignments**



UniProt data

- Download latest release**
Get the UniProt data
- Statistics**
View Swiss-Prot and TrEMBL statistics
- How to cite us**
The UniProt Consortium

Protein spotlight

Out Of The Ordinary

January 2017

Life depends on chemical signals. Without them our heart wouldn't know how to beat or our thoughts how to form, our eyes would be unable to see, our legs unable to walk and our mouths would be incapable of speech...

всесторонний, высококачественный и свободно доступный ресурс по последовательностям белков и информации о их функциях.

<http://www.uniprot.org>

Welcome

Deposit

Search

Visualize

Analyze

Download

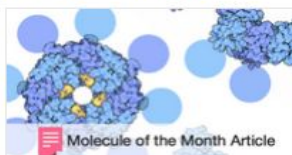
Learn

A Structural View of Biology

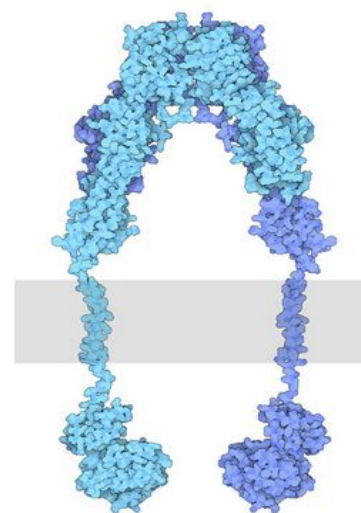
This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Structure and Health Focus: Ebola



February Molecule of the Month



Insulin Receptor

PDB – Protein data bank. База данных содержащая 3D структуры отдельных макромолекул и их комплексов (в том числе с низкомолекулярными соединениями).

<http://www.pdb.org>



InterPro
Protein sequence analysis & classification

Examples: IPR020405, kinase, P51587, PF02932, GO:0007165

[Home](#)[Search](#)[Release notes](#)[Download](#)[About InterPro](#)[Help](#)[Contact](#)

InterPro: protein sequence analysis & classification

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. We combine protein signatures from a number of member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool. [Read more about InterPro](#)

Analyse your protein sequence



InterPro 49.0
20th November
2014

Features include:

- An update to PROSITE patterns (20.105), PROSITE profiles (20.105)
- Integration of 817 new methods from the PANTHER, PROSITE profiles, Pfam and SUPERFAMILY databases.

[Read more](#)

InterPro классифицирует последовательности на уровне суперсемейств, семейств и подсемейств, предсказывает возможные функциональные домены, повторы и важные сайты. Предсказывает белковые сигнатуры "signatures" используя классификации и механизм автоматической аннотации белков и геномов.

<http://www.ebi.ac.uk/interpro>

Крупнейшие научные проекты-генераторы биомедицинских данных

- * **Human proteom project:** Human Protein Atlas <http://www.proteinatlas.org/>
- * **1000 Genomes** <http://www.1000genomes.org/>
- * **ENCODE (ENCyclopedia Of DNA Elements)** <http://www.genome.gov/encode/>
- * **Genome-wide association studies (GWAS):** <http://www.ebi.ac.uk/gwas/>
- * **COSMIC (Catalogue Of Somatic Mutations In Cancer)**
<http://cancer.sanger.ac.uk/cosmic/>
- * **Human Microbiom project** <http://www.hmpdacc.org/>

THE HUMAN PROTEIN ATLAS



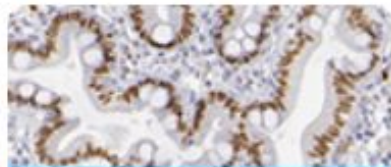
ABOUT & HELP



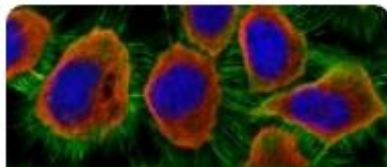
A Tissue-Based Map of the Human Proteome

Here, we summarize our current knowledge regarding the human proteome mainly achieved through antibody-based methods combined with transcriptomics analysis across all major tissues and organs of the human body. A large number of lists can be accessed with direct links to gene-specific images of the corresponding proteins in the different tissues and organs.

[Read more](#)



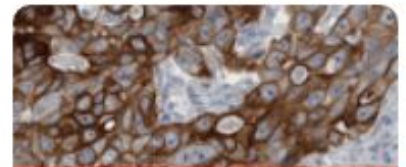
TISSUE ATLAS



SUBCELL ATLAS



CELL LINE ATLAS



CANCER ATLAS

SEARCH ? »

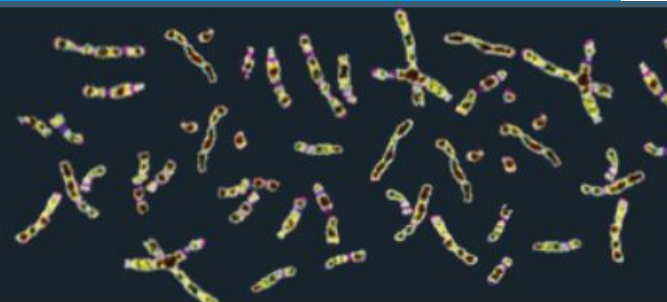
e.g. insulin, PGR, CD36

Search

[Fields »](#)

1000 Genomes

A Deep Catalog of Human Genetic Variation



[Home](#) [About](#) [Data](#) [Analysis](#) [Participants](#) [Contact](#) [Browser](#) [Wiki](#) [FTP search](#)

LATEST ANNOUNCEMENTS

WEDNESDAY SEPTEMBER 30, 2015

A global reference for human genetic variation

The Phase 3 publication, [A global reference for human genetic variation](#) and the Phase 3 Structural variation publication, [An integrated map of structural variation in 2,504 human genomes](#) are now available from [Nature](#) alongside a [celebration of 25 years of the Human Genome Project](#)

The variants from the Phase 3 analysis are available in <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/> and extended information about the SV dataset can be found in ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/phase3/integrated_sv_map/.

Both these papers are open access and should be free for everyone to read and download.

If you have any questions about the data these papers are based on or how to access it please email info@1000genomes.org

<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>

Recent project announcements

WEDNESDAY DECEMBER 16, 2015

GRCh38 alignments for Exome and High Coverage 1000 Genomes Data

We have realigned exome data from 2692 samples and high coverage PCR-free data from 24 samples, generated for the 1000 Genomes

NAVIGATION

- [Frequently Asked Questions](#)

LINKS



[All Project Announcements](#)



[Sample and Project Information](#)



[Media Archive](#)



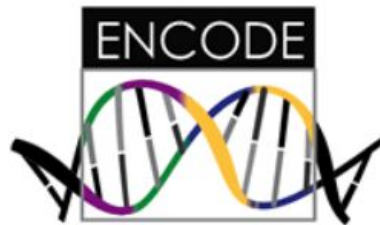
[Find the 1000 Genomes Project Publications](#)



Division of Genome Sciences

[Centers of Excellence in Genomic Science](#) ▶[Division Staff](#) ▶[ENCODE Project](#) ▶[Functional Analysis Program](#)[Genetic Variation Program](#) ▶[Genome Informatics and Computational Biology Program](#) ▶[Genome Technology Program](#) ▶[NHGRI Genome Sequencing Program \(GSP\)](#) ▶

The ENCODE Project: ENCyclopedia Of DNA Elements

[+ Share](#) [Print](#)

- ↓ [Overview](#)
- ↓ [Publications, Features and Press Releases](#)
- ↓ [Consortium Membership](#)
- ↓ [Data Release Policy](#)
- ↓ [Accessing ENCODE Data](#)
- ↓ [ENCODE Tutorials](#)
- ↓ [Common Cell Types](#)
- ↓ [Requests for Application \(RFAs\)](#)
- ↓ [Program Staff](#)


Follow the ENCODE Project on:

[Facebook](#)[Twitter](#)

ENCODE Overview

The National Human Genome Research Institute (NHGRI) launched a public research consortium named ENCODE, the **Encyclopedia Of DNA Elements**, in September 2003, to carry out a project to identify all functional elements in the human genome sequence. The project started with two components - a pilot phase and a technology development phase.

See Also:

[Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project](#) 
Nature, June 13, 2007

[Major Findings from The ENCODE Pilot Project](#)
June 2007

[The modENCODE Project](#)

[Grants Home](#)

On Other Sites:

[The ENCODE \(ENCyclopedia Of DNA Elements\) Project](#).
Science, October 22, 2004

[Nature ENCODE explorer](#)
Sept. 5, 2012

ENCODE Resources:

[ENCODE Web Focus](#)
Related articles on ENCODE from *Nature*

Genome-wide association studies (GWAS)

GWAS / Diagram

This diagram shows all SNP-trait associations with $p\text{-value} \leq 5.0 \times 10^{-8}$, published in the GWAS Catalog

Filter the
diagram

Filter by ti

Clear

Apply

Show SNPs for

343
Digestive
system disease

184
Cardiovascular
disease

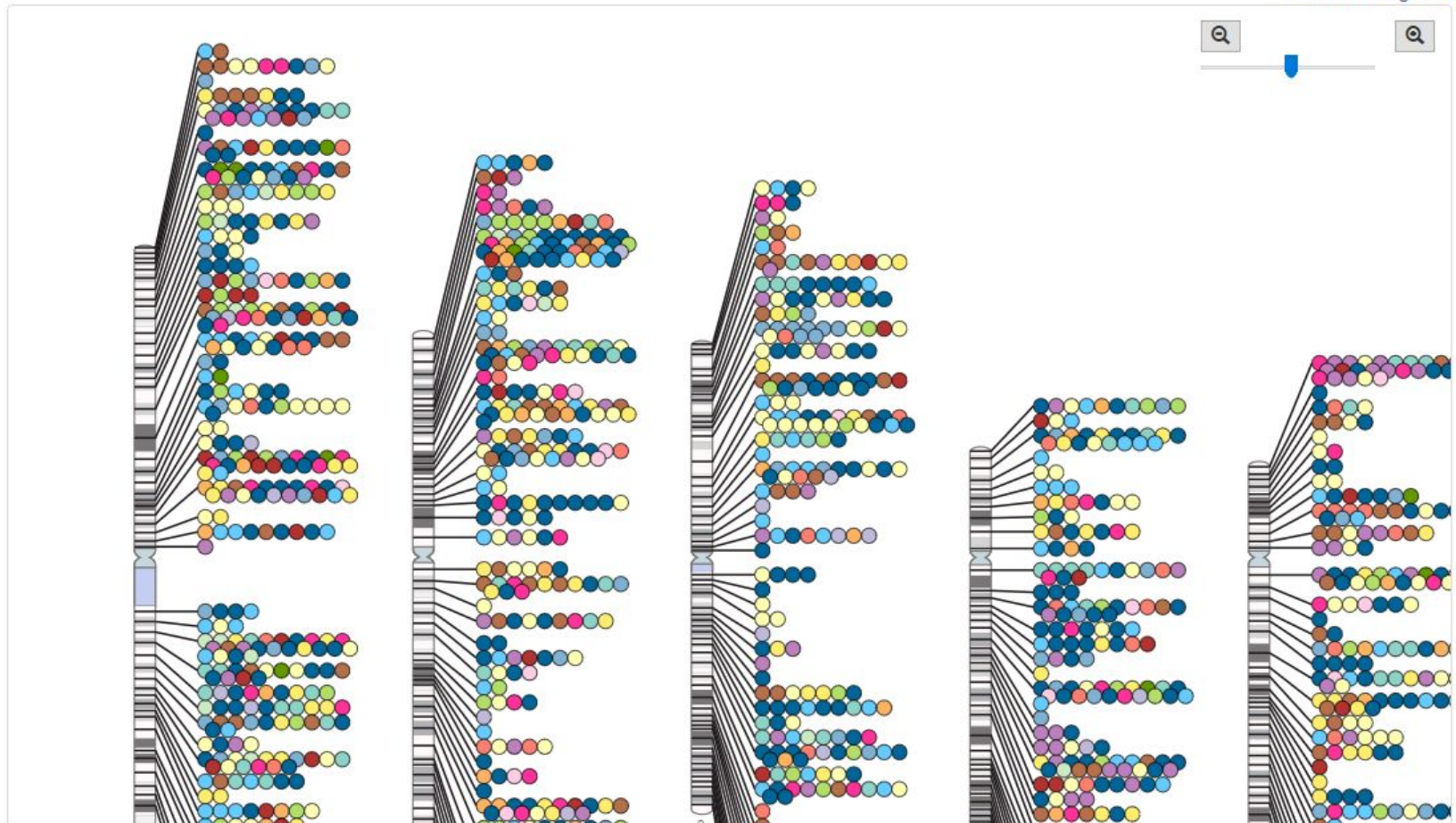
181
Metabolic
disease

330
Immune
system disease

517
Nervous
system disease

62
Liver
enzyme
measurement

Download diagram



COSMIC v75

eg: Braf, COLO-829, Carcinoma, V600E, BRCA-UK, Campbell

SEARCH

R Resources

Key COSMIC resources

[Cell Lines Project](#)
[COSMIC Whole Genomes](#)
[Cancer Gene Census](#)
[Drug Sensitivity](#)
[Mutational Signatures](#)
[GRCh37 Cancer Archive](#)

T Tools

Additional tools to explore COSMIC

[Cancer Browser](#)
[Genome Browser](#)
[CONAN](#)
[Beacon^{New}](#)
[COSMIC Mart](#)

C Expert Curation

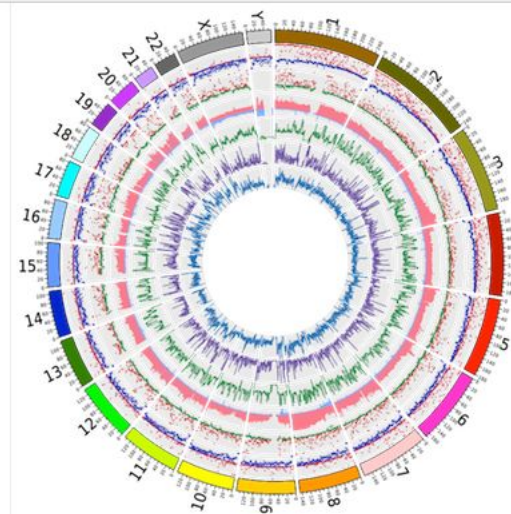
High quality curation by expert postdoctoral scientists

[Cancer Gene Census](#)
[Curated Genes](#)
[Gene Fusions](#)
[Genome-Wide Screens](#)

D Data

Further details on using COSMIC's content

[Downloads](#)
[License](#)
[Submission](#)
[How Do We Annotate Genomes?](#)
[Datasheet V75](#)



Genomic Landscape of Cancer

Cancer Gene Census



We welcome Dr. Zbyslaw Sondka, COSMIC's new dedicated [Cancer Gene Census](#) curator. Dr. Sondka is challenged with expanding the Census with new genes from literature publications and cancer genome studies, and describing in much more detail which diseases each gene causes, via which mutations and mechanisms, and by what precise evidence.

We expect to begin releasing enhanced Census



NIH HUMAN MICROBIOME PROJECT

Current News

- January 2015
Metagenome Analysis Workshop
March 3-6
- September 2014
IHMC 2015 from Mar. 31 to Apr. 2
- May 2014
Poster and Booth at ASM 2014

[More News Items](#)

Publications

- Microbial community assembly and metabolic function during mammalian c...
- HIV-induced immunosuppression is associated with colonization of the p...
- Comparative metabolomics in vegans and omnivores reveal constraints on...

[More Publications](#)

[Login](#)

[OVERVIEW](#)

[REFERENCE GENOMES](#)

[MICROBIOME ANALYSIS](#)

[HEALTH & ETHICS](#)

[RESOURCES](#)

[OUTREACH](#)

[DATA BROWSER](#)

[Feedback](#)

Welcome to the Data Analysis and Coordination Center (DACC) for the National Institutes of Health (NIH) Common Fund supported Human Microbiome Project (HMP). This site is the central repository for all HMP data. The aim of the HMP is to characterize microbial communities found at multiple human body sites and to look for correlations between changes in the microbiome and human health. More information can be found in the menus above and on the [NIH Common Fund site](#).

[GET DATA](#)

[GET TOOLS](#)

Areas of Interest

Human Microbial Sampling

16S RNA and whole metagenome sequencing of samples collected from 300 healthy human participants, to characterize complexity of microbial communities at individual body sites and to provide insights into functions performed by the human microbiome...

[DACC Member Organizations](#)

Полезные инструменты и ресурсы

FireFox toolbar (BioFox, BioBar)

<http://www.bioinformatics.org>

<http://www.netsci.org/Resources/Software/index.html>

**[http://en.wikipedia.org/wiki/Category:
Bioinformatics_software](http://en.wikipedia.org/wiki/Category:Bioinformatics_software)**

<http://www.bioinformatics.ru>

**Software and Databases for Computational Biology on the Internet:
<http://www.cbcb.umd.edu/~salzberg/appendixa.html>**

sci-hub

Литература:

- * **J. Pevsner. Bioinformatics and functional genomics. 2015.**
- * А.Леск **Введение в биоинформатику**
- * Ж. Сетубал, Ж. Мейданис **Введение в вычислительную молекулярную биологию**
- * Э. МакКонки **Геном человека**
- * У. Клаг, М. Каммингс **Основы генетики**
- * Б. Нолтинг **Новейшие методы исследования биосистем**
- * И. Сарвилина, В. Каркищенко, Ю. Горшкова **Междисциплинарные исследования в медицине**
- * М.А.Каменская **Информационная биология**
- * С.Примроуз, Р.Тваймен **Геномика. Роль в медицине**
- * Р.Дурбин, Ш.Эдди, А.Крог., Г.Митчисон **Анализ биологических последовательностей**
- * М.Бордовский, С.Екишева **Задачи и решения по анализу биологических последовательностей**
- * Б.Хаубольд, Т.Вие **Введение в вычислительную биологию. Эволюционный подход**