

УЧЕБНИК

**Н.Ю. Часовских**

# БИОИНФОРМАТИКА

Министерство науки и высшего образования РФ

Рекомендовано Координационным советом по области образования «Здравоохранение и медицинские науки» в качестве учебника для использования в образовательных учреждениях, реализующих основные профессиональные образовательные программы высшего образования по направлениям подготовки 30.05.01 «Медицинская биохимия», 30.05.02 «Медицинская биофизика», 30.05.03 «Медицинская кибернетика»

Регистрационный номер рецензии 783 от 20 июня 2019 года



**Москва**  
ИЗДАТЕЛЬСКАЯ ГРУППА  
«ГЭОТАР-Медиа»  
2020

## ОГЛАВЛЕНИЕ

Список сокращений .....	5
Введение .....	6
<b>Глава 1.</b> Классификация баз данных в биоинформатике .....	9
<b>Глава 2.</b> NCBI .....	21
2.1. Наиболее популярные ресурсы NCBI .....	23
2.2. Руководство к NCBI и система Entrez .....	28
2.3. Работа с NCBI .....	33
<b>Глава 3.</b> EMBL-EBI .....	46
3.1. Характеристика EMBL-EBI .....	46
3.2. Работа с EMBL-EBI .....	48
<b>Глава 4.</b> Геномный браузер ENSEMBL .....	52
4.1. Характеристика Ensembl .....	52
4.2. Аннотация генов Ensembl .....	57
4.3. Геномный браузер Ensembl .....	61
4.4. Работа с геномным браузером Ensembl .....	63
<b>Глава 5.</b> Геномный браузер UCSC .....	72
5.1. Характеристика UCSC .....	72
5.2. Работа с UCSC .....	77
<b>Глава 6.</b> Парное выравнивание последовательностей .....	84
6.1. Выполнение глобального и локального парного выравнивания .....	90
<b>Глава 7.</b> BLAST .....	98
7.1. Характеристика BLAST .....	98
7.2. Работа с BLAST .....	113
<b>Глава 8.</b> Множественное выравнивание последовательностей .....	119
8.1. Определение, этапы и методы множественного выравнивания .....	119
8.2. Выполнение множественного выравнивания .....	129
<b>Глава 9.</b> Молекулярная эволюция, филогения .....	149
9.1. Определение и гипотезы молекулярной эволюции .....	149
9.2. Филогенетический анализ .....	150
9.3. Работа с MEGA .....	153

<b>Глава 10.</b> Анализ экспрессии генов (микрочипы) . . . . .	166
10.1. Анализ экспрессии генов с помощью микрочипов. . . . .	166
10.2. Работа с данными экспериментов Microarray в GEO и GEO2R. . . . .	169
<b>Глава 11.</b> Анализ биологических путей . . . . .	190
11.1. Базы данных биологических путей . . . . .	190
11.2. Работа с базами данных биологических путей. . . . .	193
<b>Глава 12.</b> Изучение структуры и функций белков . . . . .	206
12.1. Методы изучения свойств белков . . . . .	206
12.2. PDB . . . . .	210
12.3. UniProt . . . . .	228
12.4. Работа с инструментами для изучения свойств белков . . . . .	232
<b>Глава 13.</b> Молекулярный докинг . . . . .	257
13.1. Определение, оценочные функции, программы для молекулярного докинга . . . . .	257
13.2. Молекулярное моделирование взаимодействий белка и лиганда с помощью AutoDockTools . . . . .	262
<b>Глава 14.</b> Предсказание структуры и функций белков . . . . .	271
14.1. Методы предсказания свойств белков. . . . .	271
14.2. Функциональный анализ и определение структуры алкалинпротеазы. . . . .	275
<b>Глава 15.</b> Анализ метаболизма химических соединений . . . . .	299
<b>Глава 16.</b> Биотехнические системы и биоинформационные ресурсы. . . . .	322
Заключение. . . . .	341
Рекомендуемые литература и интернет-ресурсы . . . . .	344

## Глава 2

### NCBI

Важную роль в реализации биоинформационных исследований на сегодня играет National Center for Biotechnological Information, NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), разрабатывающий новые информационные технологии для исследования молекулярно-генетических процессов. Это создание систем хранения и анализа биологической информации, передовые технологии обработки информации, облегчение доступа к БД и программному обеспечению, координация проектов по сбору биотехнологической информации в мировом масштабе.

NCBI обслуживает БД последовательностей DNA GenBank: создаются записи о структуре расшифрованных последовательностей (данные научных публикаций, информация от исследователей) и осуществляется обмен ими с EMBL — European Molecular Biology Laboratory (<http://www.embl.org/>) и DDBJ DNA — Data Bank of Japan (<http://www.ddbj.nig.ac.jp/index-e.html>).

Информационно-поисковая система NCBI интегрирует информацию, содержащуюся во всех БД, — DNA, RNA, белковых последовательностей и структур, геномов, таксономии, библиографии и других, а также содержит различные стандартные программы биоинформатики (например, BLAST).

#### *Ключевые элементы NCBI*

##### 1. PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>)

Раздел научной литературы, содержит поисковый сервис Национальной медицинской библиотеки (National Library of Medicine, USA). Поддерживает online-связь с журналами, на сайте располагается руководство PubMed; каждой статье присваивается уникальный номер PMID.

##### 2. GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>)

БД генетических последовательностей, аннотированная коллекция всех общедоступных последовательностей DNA.

##### 3. RefSeq (<http://www.ncbi.nlm.nih.gov/refseq/>)

БД референсных последовательностей (Reference Sequences) — важнейший элемент NCBI. Содержит неповторяющиеся последова-

тельности геномных DNA, mRNA и белков, каждая из которых имеет уникальный номер.

Идентификаторы RefSeq включают различные форматы, которые делятся на два типа в зависимости от способа аннотации:

- ▶ модели RefSeq — RNA и белковые продукты, которые генерируются при конвейерном подходе к аннотации генома эукариот, в этих записях используются префиксы XM\_, XR\_ и XP\_;
- ▶ известные RefSeq — для RNA и белковых продуктов, которые получены из данных GenBank cDNA и EST и подтверждены курирующей группой RefSeq.

В этих записях используются префиксы NM\_, NR\_ и NP\_ (перечень приведен ниже).

Префикс	Тип молекул	Комментарии
NC_	Genomic	Полные геномные молекулы
NG_	Genomic	Неполные геномные регионы
NT_	Genomic	Контиг
NM_	mRNA, DNA	—
NR_	RNA	—
XM_	mRNA	Модель
XR_	RNA	Модель
NP_	Protein	Ассоциирован с NM_ или NC_
XP_	Protein	Модель, ассоциирована с XM_

Информация о последовательности может быть представлена пользователю в разных видах: GenBank, GenPept, FASTA, Graphics и других, выбрать вариант можно в левом верхнем углу записи.

GenBank и GenPept содержат, помимо самой последовательности, ее различные характеристики — номера, длину, ссылки на публикации, комментарии, организм-источник, описание белка, регионов, сайтов, CDS (кодирующую последовательность).

### Формат FASTA

FASTA — очень компактный формат записи последовательности, со строкой-заголовком и строкой-последовательностью нуклеотидов или аминокислот. Он универсален, используется для работы как программ, так и людей (при открытии текстовым редактором). Допускается хранение в одном файле формата FASTA многих последовательностей. Пример записи приведен на рис. 2.1.

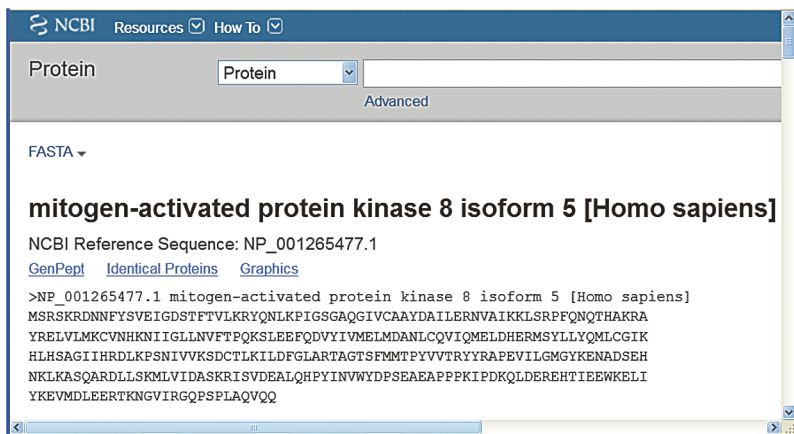


Рис. 2.1. Запись в формате FASTA

Символ > означает начало информации о последовательности. До первого пробела идет идентификатор последовательности (NP\_001265477.1 в приведенном примере). Оставшаяся информация в данной строке — ее текстовое описание (mitogen-activated protein kinase 8 isoform 5 [Homo sapiens]). Остальные строки — сама последовательность.

## 2.1. НАИБОЛЕЕ ПОПУЛЯРНЫЕ РЕСУРСЫ NCBI

ClinVar ([www.ncbi.nlm.nih.gov/clinvar/](http://www.ncbi.nlm.nih.gov/clinvar/)) является новым ресурсом медицинской генетики, который собирает данные о связях между вариациями последовательности и фенотипами человека. В ClinVar может представляться информация по изменчивости, фенотипам, интерпретации медицинского значения вариаций, а также сведения об источнике информации. ClinVar интегрирует данные из различных источников и добавляет значения, такие как номенклатура HGVS, идентификаторы dbSNP или dbVar.

ClinVar также сопоставляет наименования представленных фенотипов в понятиях, которые интегрированы в MedGen. Помимо самой информации, ClinVar также сообщает уровень доверия к утверждению в зависимости от количества и типа материала. Каждое из отдельных утверждений, представленных ClinVar, имеет уникальный номер в формате SCV000000000.0, и материалы, которые связаны с тем же вариантом и фенотипом, сохраняются в записях с номером RCV000000000.0.

MedGen ([www.ncbi.nlm.nih.gov/medgen/](http://www.ncbi.nlm.nih.gov/medgen/)) — портал, который собирает информацию о человеческих заболеваниях, имеющих генетический компонент. Выпущенный в 2012 г., MedGen превратился в важный узел NCB, который стандартизирует представление фенотипов человека и поддерживает связь между БД и инструментами, зависящими от этих стандартов (например, GTR, ClinVar, dbGaP и PheGenI).

MedGen соотносит информацию о фенотипах с постоянными идентификаторами, привязывая термины, используемые для названия нарушений, к их клиническим особенностям. Когда это возможно, используется тот же идентификатор, что и в Единой медицинской языковой системе (Unified Medical Language System, [www.nlm.nih.gov/research/umls/](http://www.nlm.nih.gov/research/umls/)). MedGen использует комбинацию автоматической обработки и курирования, чтобы агрегировать эти данные, и представляет результаты в виде текстового отчета с несколькими разделами. Эти разделы могут включать (в зависимости от имеющихся данных) описание заболевания и его клинические особенности, коллекции соответствующих профессиональных руководств, клинических исследований и систематических обзоров. Отчеты также могут содержать метаданные и идентификаторы из других БД или онтологий, таких как «Онтологии фенотипов человека» (Human Phenotype Ontology, [www.human-phenotype-ontology.org/](http://www.human-phenotype-ontology.org/)), OMIM ([www.omim.org](http://www.omim.org)) и SNOMED CT ([www.ihtsdo.org/snomed-ct/](http://www.ihtsdo.org/snomed-ct/)). Кроме того, MedGen предоставляет доступ к генетическим тестам в Реестре генетического тестирования (NIH Genetic Testing Registry, GTR) и соответствующим записям других ресурсов NCBI, в том числе PubMed, Gene и ClinVar. MedGen поддерживает запросы по терминам, идентификаторам БД, таких как MIM или SNOMED CT, и принципиальным связям, таким как нарушения обмена и клиническая характеристика. В дополнение к веб-службе MedGen доступен через программные утилиты Entrez Programming Utilities (E-утилиты) и FTP (<ftp://ftp.ncbi.nlm.nih.gov/pub/medgen/>).

PubReader ([www.ncbi.nlm.nih.gov/pmc/about/pubreader/](http://www.ncbi.nlm.nih.gov/pmc/about/pubreader/)) — удобный вариант отображения для просмотра полного текста статьи в PubMed Central (PMC). Новый формат, названный PubReader, использует возможности HTML5 и CSS3 для решения проблем, возникающих при чтении научных статей на планшетах и других устройствах с небольшим экраном, но также хорошо работает на ноутбуках и настольных компьютерах.

PubReader может отображать любые статьи PMC, доступные в полнотекстовом формате HTML, и автоматически разбивает текст на несколь-

ко столбцов и страниц в зависимости от размера экрана пользователя. Полоса изображений появляется в нижней части дисплея и отображает миниатюры любых рисунков и таблиц в документе. При нажатии на миниатюру она открывается в увеличенном виде контента, и изображение остается на месте, так что такой доступ возможен с любой страницы в статье. Это позволяет просматривать любой рисунок или таблицу, не теряя место чтения в статье.

SciENcv (<https://www.ncbi.nlm.nih.gov/account/>) — сеть автобиографий научных экспертов Science Experts Network Curriculum Vitae (SciENcv), новый инструмент в My NCBI. SciENcv предназначен для ученых, которые или претендуют на получение научно-исследовательских инвестиций NIH, или связаны с ними. SciENcv помогает пользователям сформировать и поддерживать профиль для создания биографического очерка и его использования в грантовых заявках.

Основной целью системы является снижение административного бремени для исследователей и учреждений в ходе работы с грантами за счет оптимизации задачи создания биографического очерка и использования данных, которые могут существовать как личный профиль резюме (Personal Profile Summary). SciENcv включает в себя данные из «Моей библиографии» (My Bibliography), сервиса, который финансируемые NIH ученые используют для управления сведениями о публикациях, что интегрируется с рядом других БД, таких как PubMed, PMC, NIH Manuscript Submission System (NIHMS), SPIRES и eRA Commons. Пользователи имеют возможность сохранить свой профиль закрытым или сделать его общедоступным и поделиться им с коллегами.

Gene (<https://www.ncbi.nlm.nih.gov/gene/>) интегрирует информацию о генах по широкому спектру видов от вирусов, бактерий до эукариот. Запись может содержать номенклатуру, RefSeq-последовательности, карты, фенотипы, переходы к распространенным геном-фенотип-локус-специфическим ресурсам.

При выводе данных на экран Display Settings (рис. 2.2) пользователь может изменять их конфигурацию, при этом возможны два типа представления информации: ранжированные результаты запроса (вверху) или дисплей с определенной записью (внизу).

PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) содержит библиографические описания и резюме из областей биомедицины и здравоохранения, биоинформатики, химических наук и биоинженерии.



The screenshot displays the NCBI Gene database interface. The top section shows a search for 'Gene: stx2[sym]' with options to 'Save search' or 'Advanced'. Below this, a 'Display Settings' dropdown is set to 'Tabular, 20 per page, Sorted by Gene Weight'. A table of search results is visible, listing gene names, IDs, and chromosome locations. A 'Format' dropdown menu is open, showing options like 'Summary', 'Tabular', 'ASN 1', 'XML', and 'UI List'. The second part of the screenshot shows the 'Full Report' for 'Stx2' in 'Homo sapiens (human)'. It includes details such as 'Organism: Homo sapiens', 'Lineage: Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi, Mammalia, Eutheria, Euarchontoglires, Primates, Haplorhini, Catarrhini, Hominidae, Homo', and 'Also known as: EPM, EPM, STX2A, STX2B, STX2C'. A summary states: 'The product of this gene belongs to the syntaxin/epimorphin family of proteins. The syntaxins are a large protein family'.

Рис. 2.2. Вывод данных в Gene

В PubMed также предоставлен доступ к дополнительным сайтам и ссылки на медико-биологические ресурсы NCBI. Кроме того, сейчас в резюме публикаций включаются ключевые слова авторов на других (помимо английского) языках, если авторы предоставили соответствующие данные. Дополнительный выбор языка доступен по ссылкам на дисплее *Abstract*, где жирным выделением текста указывается язык, используемый в данный момент. Интерфейс поиска в PubMed постоянно совершенствуется.

На странице расширенного поиска пользователи могут скачать свои истории поиска в файл CSV, что позволяет сохранить запись комплекса поисков.

**BLAST** рассмотрен в главе 7.

Genome (<https://www.ncbi.nlm.nih.gov/genome>) — БД последовательностей и карт, аннотаций целых геномов свыше 1000 видов и штаммов. Все три основных домена дерева жизни (бактерии, археи, эукариоты), представленные вирусами, фагами, вириоидами, плазмидами, органеллами, присутствуют на данном ресурсе.

**RefSeq** — коллекция, содержащая всеобъемлющий, комплексный, недублированный, хорошо аннотированный набор последовательностей, в том числе геномной DNA, транскриптов и белков. Последовательности RefSeq используют в качестве основы для исследований в медицине. RefSeq предоставляет постоянные ссылки для аннотации генома, идентификации и описания генов, мутаций и анализа полиморфизмов (особенно RefSeqGene-записей), исследования экспрессии и сравнительного анализа.

**Taxonomy** является курируемой классификацией и номенклатурой для всех организмов в публичной БД последовательностей. К текущему моменту это составляет около 10% всех описанных видов планеты.

Браузер Taxonomy предлагает два различных типа веб-страниц:

- 1) страница иерархии, которая представляет ступени-семейства таксономической классификации (рис. 2.3);
- 2) таксон-специфичная страница, которая суммирует всю информацию, ассоциированную с какой-либо определенной таксономической записью в БД (рис. 2.4).

Пример отображения таксономии по первому типу — браузер с иерархическим отображением семейства Hominidae.

Здесь отображены четыре общих рода (*Gorilla*, *Homo*, *Pan*, *Pongo*) с уровнями-видами (*Gorilla gorilla*, *Homo sapiens*, *Pan paniscus*, *Pan troglodytes*, *Pongo pygmaeus*, *Pongo sp.*) и двумя подвидами.

Общие названия приведены в скобках, если они имеются в БД таксономии. Происхождение гоминид отображается выше в строке в верхней части дисплея. Выбрав любой из таксонов выше Hominidae (в родословной) или ниже (в иерархическом дисплее), можно переориентировать браузер на этот таксон. Выбрав Hominidae, можно будет отображать страницу таксонов для гоминид.

На таксон-специфичном дисплее перечислены все имена, связанные с определенным таксоном (наименование, ранг, генетические коды и др.). Показана полная родословная; выбирая слово с переходом (подчеркнуто), можно переключиться с сокращенной на полную версию. Там также может быть ссылка на источник и комментарии с гиперссылками на соответствующие источники. Справа отображается число

(a) NCBI Taxonomy Browser interface. The search bar contains "As complete name" and "lock" is checked. The filter is set to "none". The display level is set to "3". The lineage is shown as: [root](#); [cellular organisms](#); [Eukaryota](#); [Fungi/Metazoa group](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Coelomata](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Primates](#); [Catarrhini](#).

(b) Hierarchy of the Hominidae family:

- [Hominidae](#) *Click on name to get more information.*
  - [Homo/Pan/Gorilla group](#)
    - [Gorilla](#)
      - [Gorilla gorilla](#) (gorilla)
    - [Homo](#)
      - [Homo sapiens](#) (human)
    - [Pan](#) (chimpanzees)
      - [Pan paniscus](#) (pygmy chimpanzee)
      - [Pan troglodytes](#) (chimpanzee)
    - [Pongo](#)
      - [Pongo pygmaeus](#) (orangutan)
        - [Pongo pygmaeus abelii](#) (Sumatran orangutan)
        - [Pongo pygmaeus pygmaeus](#) (Bornean orangutan)
      - [Pongo sp.](#)

Рис. 2.3. Страница иерархии семейства Hominidae

соответствующих записей в БД Entrez. Ссылки на внешние ресурсы отображаются в нижней части страницы.

## 2.2. РУКОВОДСТВО К NCBI И СИСТЕМА ENTREZ

Руководство к NCBI (NCBI Guide) является интерактивной директорией (<https://www.ncbi.nlm.nih.gov/guide/all/#howtos>). На главной странице NCBI Guide категории в меню ресурсов со стандартными заголовками дублируются в списке в левой части страницы (рис. 2.5). Кликнув на любую категорию, можно посмотреть список соответствующих ресурсов по четырем группам: БД (Databases), скачивание (Downloads),

The screenshot shows the NCBI Taxonomy Browser interface. At the top, there are navigation tabs for PubMed, Entrez, BLAST, OMM, Taxonomy, and Structure. A search bar is present with a search filter set to 'none' and a dropdown menu showing 'As complete name'. Below the search bar, the results for 'Chordata' are displayed.

**Chordata**

Taxonomy ID: 7711  
 Genbank common name: chordates  
 Rank: phylum  
 Genetic code: Translation table 1 (Standard)  
 Mitochondrial genetic code: Translation table 5

Other names:  
 chordates[blast name]

Lineage (full)  
 cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa;  
 Bilateria; Coelomata; Deuterostomia

Entrez records		
Database name	Subtree links	Direct links
Nucleotide	16794559	-
Protein	543310	-
Structure	8048	-
Genome	302	-
Popset	2568	5
SNP	3935062	-
3D Domains	31356	-
UniGene	365688	-
UniSTS	229501	-
PubMed Central	9924	29
Taxonomy	22428	1

**External Information Resources (NCBI LinkOut)**

LinkOut	Subject	LinkOut Provider
<a href="#">Chordata</a>	taxonomy/phylogenetic	<a href="#">Animal Diversity Web</a>
<a href="#">Fauna Iberica</a>	taxonomy/phylogenetic	<a href="#">Fauna Iberica</a>
<a href="#">ITIS</a>	taxonomy/phylogenetic	<a href="#">Integrated Taxonomic Information System</a>
<a href="#">Mikko</a>	taxonomy/phylogenetic	<a href="#">Mikko's Phylogeny Archive</a>
<a href="#">Palaeos</a>	taxonomy/phylogenetic	<a href="#">Palaeos</a>
<a href="#">ToL</a>	taxonomy/phylogenetic	<a href="#">Tree of Life</a>
<a href="#">Chordata</a>	taxonomy/phylogenetic	<a href="#">TreeBase</a>
<a href="#">UCMP</a>	taxonomy/phylogenetic	<a href="#">UCMP phylogeny exhibit</a>
<a href="#">ZooGuide</a>	taxonomy/phylogenetic	<a href="#">Zoological Record Internet Resource Guide</a>

Рис. 2.4. Таксон-специфичный вывод информации

представление сведений (Submissions) и инструменты (Tools). Список пошаговых руководств также доступен в How To на этих страницах. Популярные ресурсы отображаются внизу главной страницы NCBI Guide, а также в поле Popular Resources на главной странице NCBI.

Entrez ([www.ncbi.nlm.nih.gov/gquery/](http://www.ncbi.nlm.nih.gov/gquery/)) — интегрированная поисковая система NCBI, которая обеспечивает доступ к разнообразным наборам из нескольких десятков БД, которые содержат около 1 млрд записей.

Результаты поиска разносятся по шести основным блокам с БД с ссылками для каждой из них (рис. 2.6). Entrez поддерживает текстовый поиск, используя простые логические запросы, загрузку данных

NCBI Resources How To Sign in to NCBI

NCBI  
National Center for Biotechnology Information

All Databases 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
- Variation

**All Resources**

All Databases Downloads Submissions Tools How To

**How To**

- Save text searches and set up automated searches with E-mailed results
- Submit data to NCBI
- Find information on qvcans and qvcobiology resources
- Find bioassays in which a given drug is active
- Find bioassays that test a particular disease or protein target
- Submit sequence data to NCBI
- Download NCBI Software
- Retrieve all sequences for an organism or taxon
- Find the function of a gene or gene product
- Obtain genomic sequence for/near a gene, marker, transcript or protein
- View all SNPs associated with a gene
- Find expression patterns
- Find genes associated with a phenotype or disease
- Find human variations associated with a phenotype or disease (clinical association)
- View/download features around an object or between two objects on a chromosome
- Find sequenced genomes, including those in progress, for a taxonomic group
- Compare protein homologs between two microbial genomes
- Download the complete genome for an organism

Рис. 2.5. Категории меню руководства к NCBI

NCBI Resources How To Sign in to NCBI

Search NCBI databases Help

Search

<b>Literature</b>		<b>Genes</b>	
Books	books and reports	EST	expressed sequence tag sequences
MeSH	ontology used for PubMed indexing	Gene	collected information about gene loci
NLM Catalog	books, journals and more in the NLM Collections	GEO Data Sets	functional genomics studies
PubMed	scientific & medical abstracts/citations	GEO Profiles	gene expression and molecular abundance profiles
PubMed Central	full-text journal articles	HomoloGene	homologous gene sets for selected organisms
<b>Health</b>		PopSet	sequence sets from phylogenetic and population studies
ClinVar	human variations of clinical significance	UniGene	clusters of expressed transcripts
dbGaP	genotype/phenotype interaction studies	<b>Proteins</b>	
GTR	genetic testing registry	Conserved Domains	conserved protein domains
MedGen	medical genetics literature and links	Protein	protein sequences
OMIM	online mendelian inheritance in man	Protein Clusters	sequence similarity-based protein clusters
PubMed Health	clinical effectiveness, disease and drug reports	Structure	experimentally-determined biomolecular structures
<b>Genomes</b>		<b>Chemicals</b>	
Assembly	genome assembly information	BioSystems	molecular pathways with links to genes, proteins and chemicals
BioProject	biological projects providing data to NCBI	PubChem BioAssay	bioactivity screening studies
BioSample	descriptions of biological source materials	PubChem Compound	chemical information with structures, information and links
Clone	genomic and cDNA clones	PubChem Substance	deposited substance and chemical information
dbVar	genome structural variation studies		
Genome	genome sequencing projects by organism		
GSS	genome survey sequences		
Nucleotide	DNA and RNA sequences		
Probe	sequence-based probes and primers		
SNP	short genetic variations		
SRA	high-throughput DNA and RNA sequence read archive		
Taxonomy	taxonomic classification and nomenclature catalog		

Рис. 2.6. Поисковая система NCBI

в различных форматах и связывание записей между БД на основе утвержденных отношений.

В простейшем варианте эти ссылки могут быть перекрестными между последовательностью и тезисами статьи, в которой это опубликовано, или, например, между последовательностью белка и его кодирующей последовательностью DNA или трехмерной (3D) структурой.

Вычисленные связи между соседними записями, например на основе компьютерного сходства между последовательностями или между тезисами PubMed, обеспечивают быстрый доступ к группам связанных записей.

Наиболее популярные ссылки отображаются как обнаруженные компоненты в правой колонке результатов поиска или страницы просмотра записей, что облегчает их поиск и изучение.

Сервис LinkOut расширяет спектр ссылок на внешние ресурсы, такие как организм-специфические геномные БД.

Извлекаемые записи могут быть отображены в различных форматах и загружены отдельно или в пакетах.

В табл. 2.1 представлены разделы Entrez и их характеристика.

**Таблица 2.1.** Разделы Entrez

БД	Количество записей	Годовой рост	Описание
<b>Literature</b>			<a href="http://www.ncbi.nlm.nih.gov/home/literature.shtml">www.ncbi.nlm.nih.gov/home/literature.shtml</a>
Books	528 176	18,2%	Книги и отчеты
PubMed Central	4 066 155	11,9%	Полнотекстовые журнальные статьи
PubMed	26 413 966	4,7%	Научные и медицинские тезисы/цитирования
MeSH	265 382	2,4%	Онтологии, используемые для PubMed-индексации
NLM Catalog	1 551 801	1,4%	Индексы коллекций NLM
<b>Health</b>			<a href="http://www.ncbi.nlm.nih.gov/home/health.shtml">www.ncbi.nlm.nih.gov/home/health.shtml</a>
GTR	48 612	52,0%	Genetic Testing Registry
ClinVar	159 184	27,4%	Варианты клинического значения
PubMed Health	62 991	14,0%	Клиническая эффективность, заболевания и отчеты по исследованию лекарственных средств
dbGaP	223 662	7,6%	Исследования взаимосвязи «генотип–фенотип»
MedGen	292 341	7,1%	Медицинская литература по генетике и ссылки

Продолжение табл. 2.1

БД	Количество записей	Годовой рост	Описание
<b>Genomes</b>			<a href="http://www.ncbi.nlm.nih.gov/home/literature.shtml">www.ncbi.nlm.nih.gov/home/literature.shtml</a>
SRA	3 092 408	82,2%	Архив последовательностей DNA и RNA
Assembly	90 727	52,3%	Информация по сборке геномов
BioSample	5 224 211	43,2%	Описание материалов биологических источников
dbVar	6 147 903	37,2%	Исследования структурных вариантов геномов
BioProject	193 972	27,4%	Биологические проекты, предоставляющие данные для NCBI
Genome	16 962	25,3%	Секвенированные геномы
SNP	819 309 474	16,1%	Короткие генетические изменения
Taxonomy	1 617 350	13,3%	Таксономическая классификация и номенклатурный каталог
Nucleotide	210 148 411	5,2%	DNA- и RNA-последовательности
Clone	38 083 613	2,0%	Геномные и cDNA-клоны
GSS	39 614 616	0,6%	Исследование геномных последовательностей
Probe	32 405 018	0,1%	Зонды и праймеры
<b>Genes</b>			<a href="http://www.ncbi.nlm.nih.gov/home/genes.shtml">www.ncbi.nlm.nih.gov/home/genes.shtml</a>
GEO Profiles	128 414 055	18,1%	Профили экспрессии генов и молекулярного разнообразия
Gene	24 351 351	13,8%	Информация о локусах генов
PopSet	257 306	11,0%	Наборы последовательностей из филогенетических и популяционных исследований
EST	76 257 001	0,3%	Последовательности Expressed Sequence Tag
UniGene	6 473 284	0,0%	Кластеры экспрессируемых транскриптов
HomoloGene	141 268	0,0%	Наборы генов-гомологов для отобранных организмов
<b>Proteins</b>			<a href="http://www.ncbi.nlm.nih.gov/home/proteins.shtml">www.ncbi.nlm.nih.gov/home/proteins.shtml</a>
Protein	307 799 547	37,7%	Последовательности белков
Structure	121 463	9,2%	Экспериментально определенные структуры биомолекул
Conserved Domains	52 411	3,5%	Консервативные домены белков
Protein Clusters	820 546	0,0%	Кластеры последовательностей подобных белков

Окончание табл. 2.1

БД	Количество записей	Годовой рост	Описание
<b>Chemicals</b>			<a href="http://www.ncbi.nlm.nih.gov/home/chemicals.shtml">www.ncbi.nlm.nih.gov/home/chemicals.shtml</a>
PubChem Compound	91 679 397	50,9%	Информация о химических соединениях и их структура
PubChem Substance	223 159 019	41,8%	Информация о субстанциях
BioSystems	879 994	9,3%	Молекулярные пути с ссылками на гены, белки и химические соединения
PubChem BioAssay	1 218 668	5,6%	Скрининговые исследования биоактивности соединений

## 2.3. РАБОТА С NCBI

### Задание 1. Поиск научных публикаций в PubMed.

Поиск статей по тематике, связанной с терапией определенного заболевания.

Найдите публикации предыдущего года выпуска (с полнотекстовой версией в свободном доступе) по тематике, связанной с лечением выбранного заболевания. Переведите на английский язык название заболевания, которое будет указано в строке поиска, а также добавьте слово *therapy*.

На сайте NCBI <https://www.ncbi.nlm.nih.gov/> в разделе Popular Resources кликните на *PubMed* (рис. 2.7).

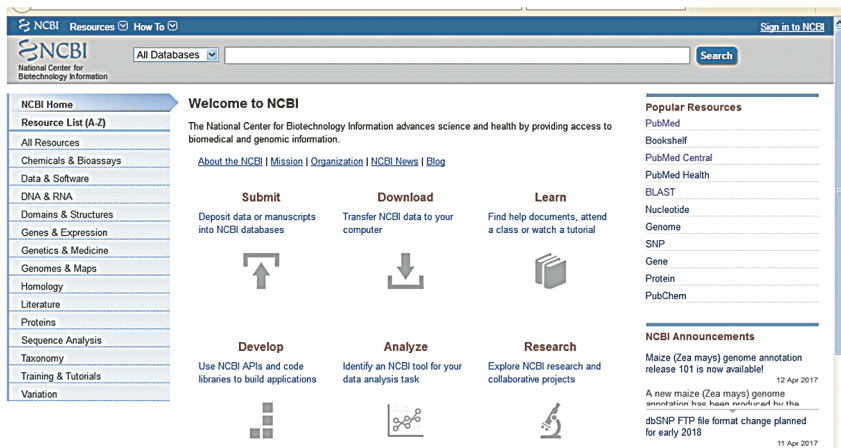


Рис. 2.7



В окне поиска наберите термин (то есть название заболевания), подтвердите запрос (рис. 2.8).

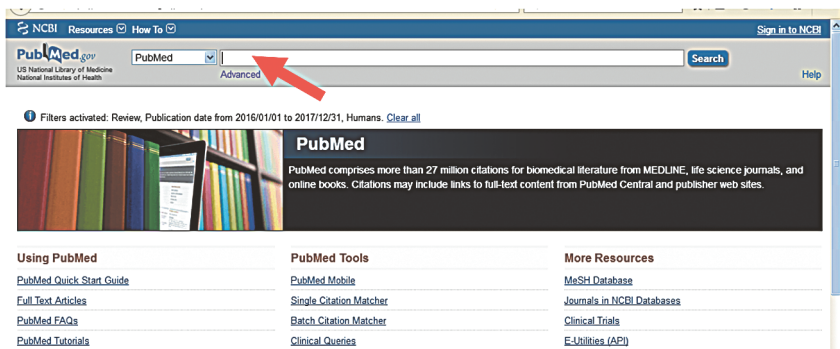


Рис. 2.8

Поскольку страницы результатов будут содержать много записей, необходимо уточнить условия поиска, для этого кликните *Advanced* (рис. 2.9).

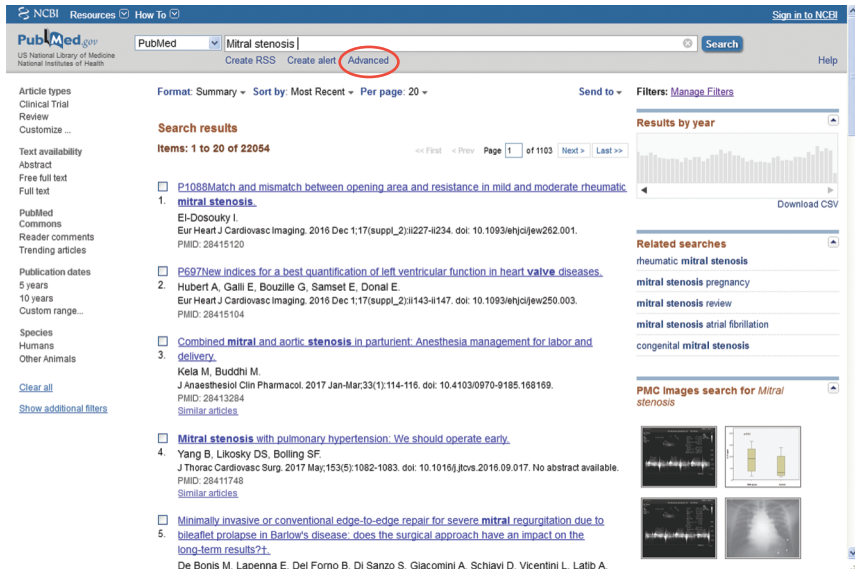


Рис. 2.9

Появится страница конструктора условий поиска *Advanced Search Builder*, на которой можно задавать дополнительные условия поиска (рис. 2.10).

Use the builder below to create your search

Edit Clear

Builder

All Fields All Fields Show index list Show index list

AND All Fields Show index list

Search or Add to history

History Download history Clear history

Search	Add to builder	Query	Items found	Time
#11	Add	Search Mitral stenosis	22054	02:07:03
#10	Add	Search Mitral stenosis Filters: Humans	20352	02:06:58
#9	Add	Search Mitral stenosis Filters: Publication date from 2016/01/01 to 2017/12/31; Humans	215	02:06:52
#8	Add	Search Mitral stenosis Filters: Review; Publication date from 2016/01/01 to 2017/12/31; Humans	23	02:06:41
#6	Add	Search Mitral stenosis therapy Filters: Review; Publication date from 2016/01/01 to 2017/12/31; Humans	21	22:35:46
#5	Add	Search Mitral stenosis therapy Filters: Clinical Trial; Review; Publication date from 2016/01/01 to 2017/12/31; Humans	26	22:35:35
#4	Add	Search Mitral stenosis therapy Filters: Clinical Trial; Publication date from 2016/01/01 to 2017/12/31; Humans	5	22:35:25
#3	Add	Search Mitral stenosis therapy Filters: Publication date from 2016/01/01 to 2017/12/31; Humans	176	22:35:02
#2	Add	Search Mitral stenosis therapy Filters: Humans	12667	22:34:42
#1	Add	Search Mitral stenosis therapy	13028	22:34:31

Рис. 2.10

Выберите вместо *All Fields* условие поиска *Title/Abstract* (поиск будет осуществляться по заголовкам и резюме публикаций) и впишите свое ключевое слово — название заболевания (рис. 2.11).

В следующей строке конструктора задайте еще одно условие — внесите *therapy* для *All Fields*, после чего запустите поиск (*Search*) (рис. 2.12).

Сократите число результатов поиска, используя опции слева (рис. 2.13).

При этом возле выбранных условий фильтрации (они выделяются синим цветом) появятся галочки (рис. 2.14).

Дополнительно ограничьте поиск публикации, задав в *Publication dates* в разделе *Custom range...* даты публикаций — введите в предложенные поля год, месяц, день, ограничивающие период последнего года, например, как показано на рис. 2.15.

Если результатов не будет, увеличьте интервал дат до 3 лет.

Resources How To

Med Home More Resources Help

**Advanced Search Builder**

Use the builder below to create your search

[Edit](#)

**Builder**

All Fields

AND

Issue

Journal

Language

Location ID

MeSH Major Topic

MeSH Subheading

MeSH Terms

Other Term

Pagination

Pharmacological Action

Publication Type

Publisher

Secondary Source ID

Subject - Personal Name

Supplementary Concept

Text word

Title

Title/Abstract

Transliterated Title

Volume

Search

History

Download history Clear history

Search	Query	Items found	Time
#11	Mitral stenosis	22054	02:07:01
#10	Mitral stenosis Filters: Humans	20359	02:06:58
#9	Mitral stenosis Filters: Publication date from 2016/01/01 to 2017/12/31; Humans	215	02:06:52
#8	Mitral stenosis Filters: Review; Publication date from 2016/01/01 to 2017/12/31; Humans	23	02:06:41
#6	Mitral stenosis therapy Filters: Review; Publication date from 2016/01/01 to 2017/12/31; Humans	21	22:35:46

Рис. 2.11

(Mitral stenosis[Title/Abstract]) AND therapy

[Edit](#) [Clear](#)

**Builder**

Title/Abstract Mitral stenosis

AND All Fields therapy

AND All Fields

Search or Add to history

**History**

Download history Clear history

Search	Add to builder	Query	Items found	Time
#11	Add	Search Mitral stenosis	22054	02:07:01
#10	Add	Search Mitral stenosis Filters: Humans	20359	02:06:58
#9	Add	Search Mitral stenosis Filters: Publication date from 2016/01/01 to 2017/12/31; Humans	215	02:06:52
#8	Add	Search Mitral stenosis Filters: Review; Publication date from 2016/01/01 to 2017/12/31; Humans	23	02:06:41
#6	Add	Search Mitral stenosis therapy Filters: Review; Publication date from 2016/01/01 to 2017/12/31; Humans	21	22:35:46
#5	Add	Search Mitral stenosis therapy Filters: Clinical Trial; Review; Publication date from 2016/01/01 to 2017/12/31; Humans	26	22:35:38
#4	Add	Search Mitral stenosis therapy Filters: Clinical Trial; Publication date from 2016/01/01 to 2017/12/31; Humans	5	22:35:28
#3	Add	Search Mitral stenosis therapy Filters: Publication date from 2016/01/01 to 2017/12/31; Humans	176	22:35:02

Рис. 2.12

PubMed (Mitral stenosis[Title/Abstract] AND therapy)

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

**Search results**

Items: 1 to 20 of 4961

1.  [Acute Myocardial Infarction Due to Coronary Artery Embolism in a 22-Year-Old Woman with Mitral Stenosis with Atrial Fibrillation Under Warfarinization: Successful Management with Anticoagulation.](#)  
Sinha SK, Jha MJ, Razi M, Chaturvedi V, Erappa Am J Case Rep. 2017 Apr 7;18:361-366. PMID: 28386054 [Similar articles](#)

2.  [Association of Transforming Growth Factor- \$\beta\$  Superfamily Class 2 Receptor-1 Gene Polymorphisms with Pulmonary Artery Hypertension Following Balloon Mitral Valvuloplasty.](#)  
Prabhu MA, Ismael S, Remani K, Nair R, Koshy J J Heart Valve Dis. 2016 Nov;25(6):708-715. PMID: 28290170 [Similar articles](#)

3.  [Conventional Surgery For Early and Late Symptomatic Mitral Valve Stenosis After MitraClip® Intervention: An Institutional Experience With Four Consecutive Patients.](#)  
Alozie A, Paranskaya L, Westphal B, Kaminski A, Steinhoff G, Sherif M, Ince H, Öner A. Heart Lung Circ. 2017 Feb 17. pii: S1443-9506(17)30028-8. doi: 10.1016/j.hlc.2016.12.008. [Epub ahead of print]

1) выберите вид **Humans**;  
2) тип статей **Article types** – **Review** (обзор);  
3) вариант представления текста:  
– **Free full text** (текст в свободном доступе)

Рис. 2.13

PubMed (Mitral stenosis[Title/Abstract] AND therapy)

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

**Search results**

Items: 1 to 20 of 40

1.  [Resolution of massive left atrial appendage thrombi with rivaroxaban before balloon mitral commissurotomy in severe mitral stenosis: A case report and literature review.](#)  
Li Y, Lin J, Peng C. Medicine (Baltimore). 2016 Dec 95(49):e5577. Review. PMID: 27930571 [Free PMC Article](#) [Similar articles](#)

2.  [The clinical challenge of concomitant aortic and mitral valve stenosis.](#)  
Unger P, Lancellotti P, de Cannière D. Acta Cardiol. 2016 Feb;71(1):3-6. doi: 10.2143/AC.71.1.3132091. Review. PMID: 26853247 [Free Article](#) [Similar articles](#)

3.  [How to define valvular atrial fibrillation?](#)  
Fauchier L, Philippart R, Clementy N, Bourguignon T, Angoulvant D, Ivanov F, Babuty D, Bernard A. Arch Cardiovasc Dis. 2015 Oct;108(10):530-9. doi: 10.1016/j.acvd.2015.06.002. Epub 2015 Jul 14. Review. PMID: 26184667 [Free Article](#) [Similar articles](#)

4.  [Catheter-based interventional strategies for cor triatriatum in the adult - feasibility study through a hybrid approach.](#)  
Li WW, Koelbgeren DR, Boums BJ, Hazekamp MG, de Mol BA, de Winter RJ. BMC Cardiovasc Disord. 2015 Jul 14;15:86. doi: 10.1186/s12872-015-0097-4. Review. PMID: 26169911 [Free PMC Article](#) [Similar articles](#)

1) выберите вид **Humans**;  
2) тип статей **Article types** – **Review** (обзор);  
3) вариант представления текста:  
– **Free full text** (текст в свободном доступе)

Рис. 2.14

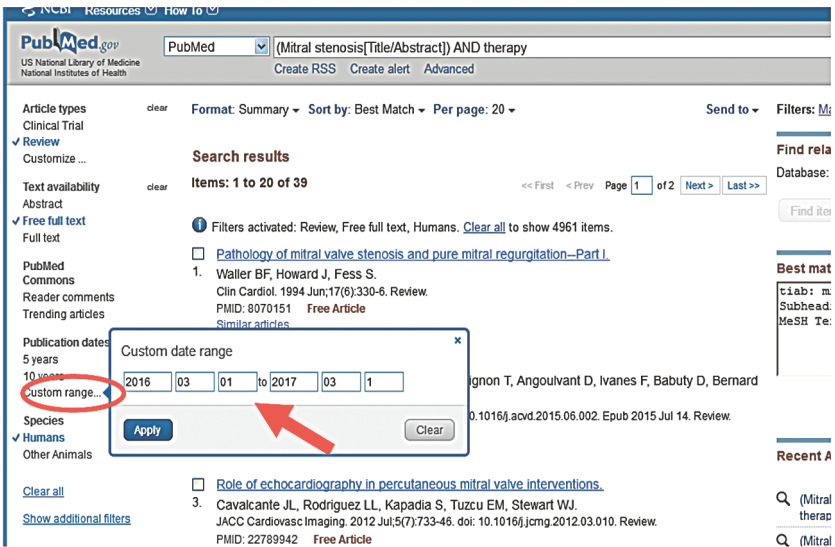


Рис. 2.15

Если список публикаций более одной страницы, в верхней части экрана выберите условие для вывода результатов *Sort by* по наилучшему совпадению *Best Match* (рис. 2.16).

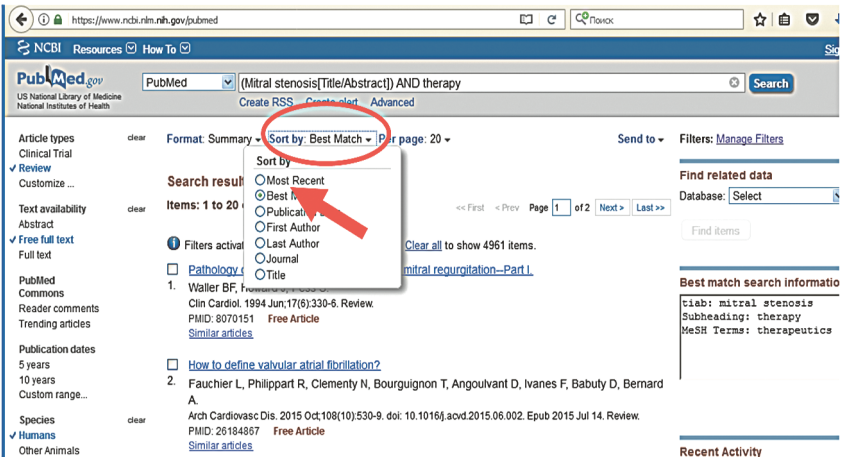


Рис. 2.16

На странице результатов кликните на одну из публикаций для просмотра *Abstract* (рис. 2.17).

The screenshot shows the PubMed search results page for the query "(Mitral stenosis[Title/Abstract]) AND therapy". The search results are sorted by "Best Match" and show 4 items. The first item is "Resolution of massive left atrial appendage thrombi with rivaroxaban before balloon mitral commissurotomy in severe mitral stenosis: A case report and literature review" by Li Y, Lin J, Peng C, published in *Medicine (Baltimore)* in 2016. The PMID is 27930571. Three red boxes highlight: 1) the PMID, 2) the "Full text links" link, and 3) the "Similar articles" link. A red arrow points to the "Full text links" link. The right sidebar shows filters and related data options.

Рис. 2.17

Внизу — индекс статьи PMID (цифра 1), справа: сверху — ссылка (-и) на полный текст *Full text links* (цифра 2), ниже — ссылки на подобные публикации *Similar articles* (цифра 3). Кликнув на название статьи, можно перейти к странице *Abstract* (рис. 2.18).

1. Кликните на *Full text links* и скачайте статью.
2. Скопируйте выходные данные публикации: журнал, год, месяц (том, номер, страницы), наименование, авторы, идентификаторы PMID и DOI.

Посмотрите информацию о первом и последнем авторах в списке, есть ли у них публикации за последние 5 лет и сколько?

Для этого кликните на фамилию первого автора, в открывшемся списке его публикаций снимите все условия, заданные вами для отбора статей (кликните на *Clear* слева) (рис. 2.19).

Далее задайте *Publication dates* — 5 years. Скопируйте фамилию автора и количество его публикаций за последние 5 лет. Найдите аналогичную информацию для последней фамилии из списка авторов, сравните

PubMed  
US National Library of Medicine  
National Institutes of Health

PubMed

Format: Abstract

Send to:

**How to define valvular atrial fibrillation?**

Faughiat L<sup>1</sup>, Philippart B<sup>2</sup>, Clementy N<sup>2</sup>, Bouquignon T<sup>2</sup>, Ancochea D<sup>2</sup>, Ibanez E<sup>2</sup>, Babuty D<sup>2</sup>, Bernard A<sup>2</sup>.

Author information

**Abstract**

Atrial fibrillation (AF) confers a substantial risk of stroke. Recent trials comparing vitamin K antagonists (VKAs) with non-vitamin K antagonist oral anticoagulants (NOACs) in AF were performed among patients with so-called "non-valvular" AF. The distinction between "valvular" and "non-valvular" AF remains a matter of debate. Currently, "valvular AF" refers to patients with mitral stenosis or artificial heart valves (and valve repair in North American guidelines only), and should be treated with VKAs. Valvular heart diseases, such as mitral regurgitation, aortic stenosis (AS) and aortic insufficiency, do not result in conditions of low flow in the left atrium, and do not apparently increase the risk of thromboembolism brought by AF. Post-hoc analyses suggest that these conditions probably do not make the thromboembolic risk less responsive to NOACs compared with most forms of "non-valvular" AF. The pathogenesis of thrombosis is probably different for blood coming into contact with a mechanical prosthetic valve compared with what occurs in most other forms of AF. This may explain the results of the only trial performed with a NOAC in patients with a mechanical prosthetic valve (only a few of whom had AF), where warfarin was more effective and safer than dabigatran. By contrast, AF in the presence of a bioprosthetic heart valve or after valve repair appears to have a risk of thromboembolism that is not markedly different from other forms of "non-valvular" AF. Obviously, we should no longer consider the classification of AF as "valvular" (or not) for the purpose of defining the aetiology of the arrhythmia, but for the determination of a different risk of thromboembolic events and the need for a specific antithrombotic strategy. As long as there is no better new term or widely accepted definition, "valvular AF" refers to patients with mitral stenosis or artificial heart valves. Patients with "non-valvular AF" may have other types of valvular heart disease. One should emphasize that "non-valvular AF" does not exclude patients with some types of valvular heart disease from therapy with NOACs.

Copyright © 2015 Elsevier Masson SAS. All rights reserved.

**KEYWORDS:** Accident vasculaire cérébral; Atrial fibrillation; Fibrillation atriale; Stroke; Valve disease; Valvulopathie

PMID: 26184887 DOI: 10.1016/j.acvd.2015.06.002

[Indexed for MEDLINE] Free full text

Publication type, MeSH terms, Substances

**Save items**

**Similar articles**

Stroke prevention strategies in patients with atrial fibrillation and heart valv [Europeas. 2016]

**Review** Non-vitamin K antagonist oral anticoagulants in atrial fibrillation [Europeas. 2016]

Efficacy and safety of novel oral anticoagulants in patients with bioprosthetic [Clin Res Cardiol. 2016]

Is early antithrombotic therapy necessary in patients with b [J Thorac Cardiovasc Surg. 2010]

**Review** Antithrombotic therapy in valvular heart disease—native and prosthetic [Chest. 2004]

**Cited by 1 PubMed Central article**

The effects of rhythm control strategies versus rate control strategies for atrial fibrillation [Syst Rev. 2017]

**Related information**

Articles frequently viewed together

MedGen

Cited in PMC

Рис. 2.18

результаты. Ответьте на вопрос, где в списке авторов размещаются ученые с наибольшим количеством публикаций.

### Задание 2. Поиск нуклеотидных последовательностей.

Найдите по поиску на NCBI <http://www.ncbi.nlm.nih.gov> нуклеотидную последовательность миоглобина и его идентификаторы для DNA, mRNA, белка: в окне запроса введите *myoglobin* для *All Databases*. После того как открылась страница результатов (рис. 2.20), перейдите в разделе *Genomes*.

*Nucleotide*, справа (*Results by taxon Top Organisms*) выберите опцию — *Homo sapiens*; если в списке нет этой позиции, необходимо кликнуть на *All other taxa*, откроется следующий список видов, в нем нужно найти *Homo sapiens* (если нет, то повторить предыдущее действие) (рис. 2.21).

Затем кликните на последовательность (автоматически откроется в формате *GenBank*):

Homo sapiens myoglobin (MB), RefSeqGene on chromosome 22  
23,591 bp linear DNA

Accession: NG\_007075.1 GI:160358355

← https://www.ncbi.nlm.nih.gov/pubmed

NCBI Resources  How To

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

PubMed

Article types: Clinical Trial, Review, Customize ...

Text availability: Abstract,  **Free full text**, Full text

PubMed Commons: Reader comments, Trending articles

Publication dates: 5 years, 10 years,  **From 2015/03/01 to 2017/03/01**

Species: Humans, Other Animals

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

Format: Summary ▾ Sort by: Author ▾ Per page: 20 ▾

### Search results

Items: 13

Filters activated: Free full text, Publication date from 2015/03/01 to :

- [Decision-Making in Clinical Practice: Oral Anticoagulant Ther...](#)  
[Atrial Fibrillation and a Single Additional Stroke Risk Factor.](#)  
Potpara TS, Dagues N, Mujović N, Vasić D, Ašanin M, Nedeljk Blomstrom-Lundqvist C, Lip GY.  
Adv Ther. 2017 Feb;34(2):357-377. doi: 10.1007/s12325-016-0458-7. E PMID: 27933569 **Free PMC Article**  
[Similar articles](#)
- [Prognosis in Familial Atrial Fibrillation.](#)  
**Fauchier L**, Bisson A, Clementy N.  
J Am Heart Assoc. 2016 Nov 19;5(11). pii: e004905. No abstract availa PMID: 27866163 **Free PMC Article**  
[Similar articles](#)
- [Erratum to '2015 HRS/EHRA/APHRS/SOLAECE expert conse...](#)  
[implantable cardioverter-defibrillator programming and testin...](#)  
[1-28\].](#)  
Wilkoff BL, **Fauchier L**, Stiles MK, Morillo CA, Al-Khatib SM, / Cuesta A, Daubert JP, Dubner S, Ellenbogen KA, Mark Estes Gasparini M, Haines DE, Healey JS, Hurtwitz JL, Keegan R, Ki Martinelli M, McGuire M, Molina LG, Okumura K, Proclemer A Teo WS, Uribe W, Viskin S, Wang CC, Zhang S.  
J Arrhythm. 2016 Oct;32(5):441-442. Epub 2016 Sep 21.  
PMID: 27761170 **Free PMC Article**  
[Similar articles](#)

Рис. 2.19



Search NCBI databases

myoglobin Search

Results found in 33 databases for "myoglobin"

Literature			Genes		
Books	263	books and reports	EST	1,100	expressed sequence tag sequences
MeSH	14	ontology used for PubMed indexing	Gene	971	collected information about gene loci
NLM Catalog	39	books, journals and more in the NLM Collections	GEO DataSets	125	functional genomics studies
PubMed	14,415	scientific & medical abstracts/citations	GEO Profiles	5,900	gene expression and molecular abundance profiles
PubMed Central	17,265	full-text journal articles	HomoloGene	2	homologous gene sets for selected organisms
Health			PopSet	184	sequence sets from phylogenetic and population studies
ClinVar	12	human variations of clinical significance	UniGene	30	clusters of expressed transcripts
dbGaP	1,509	genotype/phenotype interaction studies	Proteins		
GTR	1	genetic testing registry	Conserved Domains	6	conserved protein domains
MedGen	24	medical genetics literature and links	Protein	15,824	protein sequences
OMIM	18	online mendelian inheritance in man	Protein Clusters	3	sequence similarity-based protein clusters
PubMed Health	28	clinical effectiveness, disease and drug reports	Structure	1,200	experimentally-determined biomolecular structures
Genomes			Chemicals		
Assembly	0	genome assembly information	BioSystems	4,459	molecular pathways with links to genes, proteins and chemicals
BioProject	6	biological projects providing data to NCBI	PubChem BioAssay	18	bioactivity screening studies
BioSample	0	descriptions of biological source materials	PubChem Compound	2	chemical information with structures, information and links
Clone	857	genomic and cDNA clones	PubChem Substance	1,048	deposited substance and chemical information
dbVar	121	genome structural variation studies			
Genome	0	genome sequencing projects by organism			
GSS	4	genome survey sequences			
Nucleotide	30,843	DNA and RNA sequences			

Рис. 2.20

Сверху — основная информация о последовательности — номер, длина последовательности (23591 пара оснований), сама последовательность — DNA и др. В этом файле можно найти информацию по номеру mRNA, белка.

LOCUS NG\_00707523591 bp DNA linear PRI 04-MAY-2014

DEFINITION Homo sapiens myoglobin(MB)RefSeqGene on chromosome 22.

ACCESSION NG\_007075

The screenshot displays the NCBI Nucleotide search interface. The search term 'myoglobin' is entered in the search bar. The results are sorted by default order, showing 30843 items. The first item is 'Homo sapiens myoglobin (MB) gene, complete cds' (6,889 bp linear DNA). The right sidebar shows 'Results by taxon' with a list of organisms including Mycobacterium tuberculosis, Pseudomonas aeruginosa, and Mycobacterium abscessus.

Рис. 2.21

VERSION NG\_007075.1 GI:160358355

В разделе FEATURES информация о RNA представлена в подразделе mRNA (рис. 2.22, 2.23).

Найдите номер для RNA (NCBI Reference Sequence) transcript\_id = "NM\_203377.1".

Ниже, в подразделе CDS (кодирующая последовательность, транслируемая в белок), найдите уникальный номер белка миоглобина: protein\_id = "NP\_976311.1".

**Homo sapiens myoglobin (MB), RefSeqGene on chromosome 22**

NCBI Reference Sequence: NG\_007075.1  
[FASTA](#) [Graphics](#)

[Go to](#)

LOCUS NG\_007075 23591 bp DNA linear PRI 07-OCT-2016  
 DEFINITION Homo sapiens myoglobin (MB), RefSeqGene on chromosome 22.  
 ACCESSION NG\_007075  
 VERSION NG\_007075.1  
 KEYWORDS RefSeq; RefSeqGene.  
 SOURCE Homo sapiens (human)  
 ORGANISM [Homo sapiens](#)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
 Catarrhini; Hominidae; Homo.

COMMENT  
 REVIEWED [REFSEQ](#): This record has been curated by NCBI staff. The reference sequence was derived from [AL049747.1](#) and [AL022334.1](#). This sequence is a reference standard in the [RefSeqGene](#) project.

Summary: This gene encodes a member of the globin superfamily and is expressed in skeletal and cardiac muscles. The encoded protein is a haemoprotein contributing to intracellular oxygen storage and transcellular facilitated diffusion of oxygen. At least three alternatively spliced transcript variants encoding the same protein have been reported. [provided by RefSeq, Jul 2008].

PRIMARY	REFSEQ SPAN	PRIMARY IDENTIFIER	PRIMARY SPAN	COMP
	1-20006	AL049747.1	101-20106	c
	20007-23591	AL022334.1	111187-114771	c

FEATURES  
 Location/Qualifiers  
 source 1..23591  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /map="22q12.3"  
 5001..21591  
 /gene="MB"  
 /gene\_synonym="myoglobin; PVALB"  
 /note="myoglobin"

[gene](#)

**Customize view**

Analyze this sequence  
 Run BLAST  
 Pick Primers  
 Highlight Sequence Features  
 Find in this Sequence

**Articles about the MB gene**  
 Electroanalysis of myoglobin by electropol [Dokl Biochem Biophys]  
 A novel electrochemical aptase based on Y-[Biosens Bioelectr  
 The Distinct Gene Regulatory N Myoglobin in Prostate [PLOS O

**Reference sequence information**  
 RefSeq alternative splicing  
 See 4 reference mRNA sequen variants for the MB gene.

**More about the MB gene**  
 This gene encodes a member o globin superfamily and is expre skeletal and cardiac muscles. 1 encoded protein is a haemoprot Also Known As: PVALB, myogl

Рис. 2.22

### Задание 3. Работа с форматом FASTA.

На сайте NCBI откройте запись DNA человеческого P53 в формате FASTA по поиску Homo sapiens P53 (например, NG\_017013.2), найдите:

- 1) уникальный номер (идентификатор) последовательности;
- 2) описание последовательности;
- 3) саму последовательность.

Сохраните данную запись в виде файла.

### Вопросы для самоконтроля

1. Перечислите ключевые элементы NCBI.
2. Какие типы идентификаторов RefSeq существуют?
3. Что означают идентификаторы с префиксом NG\_, NC\_, NP\_?
4. Что означают идентификаторы с префиксом XM\_, XP\_, XR\_?
5. Что такое «формат FASTA»?
6. Перечислите наиболее популярные ресурсы NCBI.

The image shows a screenshot of the NCBI Gene database entry for myoglobin (GeneID: 4151). The main content area displays genomic features and their coordinates. The 'CDS' (Coding Sequence) feature is highlighted with a red circle. The protein coding sequence is also highlighted with a red circle. The sidebar on the right contains various links and information, including 'Homologs of the MB gene', 'Related information', and 'LinkOut to external resources'.

```

/gene="HB"
/gene_synonym="myoglobin; PVALB"
/note="myoglobin"
/db_xref="GeneID:4151"
/db_xref="HGNC:HGNC:6915"
/db_xref="MIM:160000"
mRNA
join(5001..5164,11090..11192,17249..17471,20912..21591)
/gene="HB"
/gene_synonym="myoglobin; PVALB"
/product="myoglobin, transcript variant 2"
/transcript_id="NM_203377.1"
/db_xref="GeneID:4151"
/db_xref="HGNC:HGNC:6915"
/db_xref="MIM:160000"
exon
5001..5164
/gene="HB"
/gene_synonym="myoglobin; PVALB"
/inference="alignment:Splign:2.0.8"
/number=1
exon
11090..11192
/gene="HB"
/gene_synonym="myoglobin; PVALB"
/inference="alignment:Splign:2.0.8"
/number=2
CDS
join(11098..11192,17249..17471,20912..21058)
/gene="HB"
/gene_synonym="myoglobin; PVALB"
/codon_start=1
/product="myoglobin"
/protein_id="NP_976311.1"
/db_xref="CCDS:CCDS13917.1"
/db_xref="GeneID:4151"
/db_xref="Rfam:RF0915"
/db_xref="MIM:160000"
/translation="MGLSDGEWQLVLNVGKVEADIPGHGQEVLIIRLFKGFHPETLEKF
DKFVHLKSEDEMKASEDLKKHGATVLTALGGILKKKGHEAEIKPLAQSHATKHKIPV
KYLEFISECIILQVLSKHPGDFGADACGMNKALELFRKDNASNYKELGFQG"
11132..11348
/gene="HB"
/gene_synonym="myoglobin; PVALB"
/standard_name="GDR:270294"

```

sidebar:

skeneia and carlota; muscles. encoded protein is a haemoprotein. Also Known As: PVALB, myoglin

**Homologs of the MB gene**  
The MB gene is conserved in chimpanzee, Rhesus monkey, mouse, rat, chicken, and zebra

**Related information**

Components (Core)

Full text in PMC

Gene

Map Viewer

OMIM

Protein

PubMed (Weighted)

RNA

SNP

Taxonomy

**LinkOut to external resources**

reagent reviews [ExactAntigen]

reagents [ExactAntigen]

Order MB cDNA clone/Protein

**Рис. 2.23.** Сама последовательность начинается с заголовка ORIGIN. Сохраните информацию о последовательности миоглобина в отдельном файле: справа кликните Send to, в открывшемся окне выберите Complete Record и ниже File

7. Дайте характеристику ресурсу PubMed.
8. Какие условия фильтра необходимо задать, чтобы найти в PubMed обзорные статьи за 2018 г.?
9. Какая информация содержится в PubMed на странице Abstract?
10. В каком разделе БД NCBI осуществляется поиск нуклеотидных последовательностей?