Altre basi dati di interesse disciplinare (3)

- NCBI databases
- Pharmacology free databases and additional resources
- Evidence-based Drug Resources (Clinical trials, Systematic reviews, EBM)
- Legal Databases

Lisa Dainese e Roberta Sato

May 19, 2017
Altre basi dati di interesse disciplinare (3)

- NCBI databases
- Entrez
NCBI databases

The National Center for Biotechnology Information (NCBI), a division of the National Library of Medicine (NLM) at the U.S. National Institutes of Health, is a leader in the field of bioinformatics; it studies computational approaches to fundamental questions in biology and provides online delivery of biomedical information and bioinformatics tools. NCBI hosts approximately 40 online literature and molecular biology.

The databases and resources are organized into seven concept areas: literature, genomes, variation, health, genes and gene expression, nucleotide, proteins, and small molecules and biological assays. Three additional categories encompass tools, infrastructure, and metadata.


NCBI databases

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

Entrez

Entrez is NCBI’s primary text search and retrieval system that integrates the PubMed database of biomedical literature with 38 other literature and molecular databases including DNA and protein sequence, structure, gene, genome, genetic variation and gene expression. The Entrez system comprises 39 molecular and literature databases.


Database alphabetic list and description:

For example, Entrez permits integrated searches across NCBI and NLM databases for researchers in genetic toxicology and toxicogenomics. Es. lacI gene (vedi articolo Young R.R./Toxicology 2002)
NCBI Insights
Providing insights into NCBI resources and the science behind them

RefSeq release 82 now public
Posted on May 15, 2017

RefSeq release 82 is accessible online, via FTP and through NCBI's programming utilities. This full release incorporates genomic, transcript, and protein data available as of May 8, 2017 and contains 127,098,289 records, including 84,756,971 proteins, 18,901,573 RNAs, and sequences from 69,035 organisms. The release is provided in several directories as a complete dataset and also as divided by logical groupings.

Continue reading →

Conserved Domains

The Conserved Domain database (CDD) is a collection of sequence profiles that represent highly conserved domains within protein sequences. Very often these domains have a particular function that is shared between those sequences that contain it. Typically one identifies the presence of a conserved domain in a sequence using the CD-Search tool, and these results provide access to sequence alignments, distance trees, selected literature, and structural views that highlight important elements within the domain.
Conserved Domains

The Conserved Domain Database is a resource for the annotation of functional units in proteins. Its collection of domain models includes a set curated by NCBI, which utilizes 3D structure to provide insights into sequence/structure/function relationships.

Using CDD
Quick Start Guide
How To Guides
Help
FTP
News
Publications

CDD Tools
Overview of CDD Resources
CD-Search
Batch CD-Search
CDART (Domain Architectures)
CDDTree (classification and research tool)
BLAST

Other Resources
Structure Group Home Page
Entrez Structure (Molecular Modeling Database)
Entrez Gene
Entrez Protein
BioSystems
FLink

Gene overview

Once an identifier is assigned to the concept of a gene, multiple databases connect information to that concept. Within NCBI, these databases include Gene, for primary data about the gene and portals to information about its expression, products, homologs, and phenotypes; BioSystems for pathways involving its products; GEO (see GEO chapter) and UniGene for information about expression; Bookshelf, PubMed and PubMedCentral for publications; dbGaP, PheGenI, MedGen and OMIM for phenotypes; HomoloGene for homology; dbSNP, dbVar, and ClinVar for variation; and Taxonomy for information about the organism.

- p.335 NCBI handbook, 2^ ed.
Gene

Gene is a searchable database of genes, focusing on genomes that have been completely sequenced and that have an active research community to contribute gene-specific data. Information in Gene records includes nomenclature, chromosomal localization, gene products and their attributes (e.g., protein interactions), associated markers, phenotypes, interactions, and links to citations, sequences, variation details, maps, expression reports, homologs, protein domain content, and external databases.
Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome, phenotype, and locus-specific resources worldwide.

Using Gene
Gene Quick Start
FAQ
Download/FTP
RefSeq Mailing List
Gene News
Factsheet

Gene Tools
Submit GeneRIFs
Submit Correction
Statistics
BLAST
Genome Workbench
Spilign

Other Resources
HomoloGene
OMIM
RefSeq
RefSeqGene
UniGene
Protein Clusters

Representative queries
Find genes by...
free text
human muscular dystrophy
chromosome and symbol
(lf chr OR 2 (chr)) AND adh* [sym]

HomoloGene

Examination of a gene’s function is facilitated by evaluation across multiple species, HomoloGene launched as a distinct resource in 2000, is designed to facilitate these analyses by grouping genes according to homology, providing tools for comparisons, and aggregating data for these homology groups. The HomoloGene database contains automatically generated sets of homologous genes and their corresponding mRNA, genomic, and protein sequence data from selected eukaryotic organisms. Potential homologs from other organisms are included through sequence similarity to UniGene clusters.
HomoloGene

OMIM

OMIM (Online Mendelian Inheritance in Man) is a comprehensive, authoritative compendium of human genes and genetic phenotypes. The full-text, referenced overviews in OMIM contain information on all known mendelian disorders and over 15,000 genes. OMIM focuses on the relationship between phenotype and genotype. The database allows searches of OMIM articles about human genes, genetic disorders, and other inherited traits. OMIM articles provide links to associated literature references, It is freely available and updated daily. OMIM records are hosted and served by the independent OMIM site (www.omim.org).
OMIM

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

Related Resources
- ClinVar
- Gene
- SGR
- MedGen

Last updated on: 06 May 2017
NCBI protein resources

While we can think of a single protein as a discreet polymer of amino acids, the functional form of many proteins in the cell is actually a complex of several individual polymer chains, all working in concert to do a particular task.

NCBI provides several resources that represent various aspects of proteins, ranging from the sequences of individual chains to functional classifications of large protein families. Pre-computed sets of identical and similar proteins for each sequence as determined by the BLAST algorithm are also provided. The BLAST Link (Blink) tool offers a graphical view of these pre-computed sets and provides a variety of filtering tools and links to multiple alignment views.
Protein

The Protein database is the most fundamental NCBI resource for proteins. It contains text records for individual protein sequences derived from a variety of sources.

It is important to understand that the sequences contained in almost all Protein records (with the exception of PDB - protein data bank) are conceptual translations of an RNA coding sequences, meaning that no one determined the protein sequence experimentally, but rather inferred the sequence from the corresponding RNA.

Protein records are available in several formats and are linked to many other NCBI resources, allowing users to find relevant data such as literature, DNA/RNA sequences, genes, biological pathways and expression and variation data.

- [p.385 NCBI handbook, 2^ ed.]
Protein Clusters

The Protein Clusters dataset consists of organized groups (clusters) of proteins encoded by complete and draft genomes from the NCBI Reference Sequence (RefSeq) collection of microorganisms: prokaryotes, viruses, fungi, protozoans; it also includes protein clusters from RefSeq genomes of plants, chloroplasts, and mitochondria.

The primary goal of Protein Clusters is to provide the support to functional annotation of RefSeq genomes. Functional annotation of novel proteins is based on the assumption that proteins with high levels of sequence similarity are likely to share the same function, an oversimplified model of a linear evolution where similar proteins evolve from a single ancestor further complicated by the events of gene duplication.

Protein Clusters

This collection of related protein sequences (clusters) consists of proteins derived from the annotations of whole genomes, organelles and plasmids. It currently limited to Archaea, Bacteria, Plants, Fungi, Protozoans, and Viruses.
Pubchem overview


PubChem is organized as three linked databases within the NCBI's Entrez information retrieval system. These are PubChem Substance, PubChem Compound, and PubChem BioAssay. The Substance database contains chemical informations deposited by individual data contributors** to PubChem, the Compounds database stores unique chemical structures extracted from the previous database, the BioAssay database contains biological activity data of chemical substances tested in assay experiments.

**university labs, government agencies, pharmaceutical companies, chemical vendors, publishers and a number of chemical biology resources. PubChem also hosts data from important regulatory agencies: FDA (Unique Ingrdient Identifiers-UNIIIs, pharmacological classifications). Links to patent documents (covering US, Europe and World Intellectual Property Organization since 1800) are provided from IBM, SureChem.
Corso di dottorato in Scienze Farmacologiche
Information literacy in Pharmacological Sciences 2017

PubChem
From: PubChem Substance and Compound databases
Nucleic Acids Res | Published by Oxford University Press on behalf of Nucleic Acids Research 2015. This work is written by (a) US Government employee(s) and is in the public domain in the US.

*Legenda Fig.1: Data organization in PubChem. SID, CID and AID are the identifiers for the Substance, Compound and BioAssay databases, respectively.*
Pubchem Substance

The PubChem substance database contains chemical structures, synonyms, registration IDs, description, related urls, database cross-reference links to PubMed, protein 3D structures, and biological screening results.

If the contents of a chemical sample are known, the description includes links to PubChem Compound.

PcSubstance contains more than 180 million records. You can check the count of substance records as of today. PubChem also provides a fast chemical structure similarity search tool

Pubchem Substance

The PubChem Substance Database contains descriptions of samples, from a variety of sources, and links to biological screening results that are available in PubChem BioAssay. If the chemical contents of a sample are known, the description includes links to PubChem Compound.
Pubchem Compound

The PubChem Compound Database contains validated chemical depiction information that is provided to describe substances in PubChem Substance.

Structures stored within PubChem Compound are pre-clustered and cross-referenced by identity and similarity groups. Additionally, calculated properties and descriptors are available for searching and filtering of chemical structures.

Users can perform a term/keyword search in a same manner as for substance database. In addition, the PubChem compound database also provides a chemical property search. Pccompound contains more than 63 million unique structures. You can check the count of compound records as of today https://www.ncbi.nlm.nih.gov/pccompound/
Pubchem Compound
The PubChem BioAssay Database contains BioActivity screens of chemical substances described in PubChem Substance. It provides searchable descriptions of each BioAssay, including descriptions of the conditions and readouts specific to a screening protocol.

PCBioAssay contains more than 1 million BioAssays. Each BioAssay contains a various number of data points. You can check the count of BioAssay records as of today.

PubChem BioAssay

The PubChem BioAssay Database contains bioactivity screens of chemical substances described in PubChem Substance. It provides searchable descriptions of each bioassay, including descriptions of the conditions and readouts specific to that screening procedure.
Altre basi dati di interesse disciplinare (3)

- Pharmacology free databases and additional resources
The Pharmacogenomics Knowledge Base (PharmGKB)

PharmGKB, is an interactive tool for researchers investigating how genetic variation effects drug response

https://www.pharmgkb.org/search/
Sources for PharmGKB subset of pharmagenomic data (information exchange)
http://www.futuremedicine.com/doi/full/10.2217/pgs.10.15

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genes</td>
<td>HUGO Gene Nomenclature Committee, University of California Santa Cruz Genome Browser, NCBI, Comparative Toxicogenomics Database, GeneCards, Source, MutDB, Gen Atlas, Ensembl, UniProt</td>
</tr>
<tr>
<td>Drugs</td>
<td>Drugbank, WHO, OpenEye, Comparative Toxicogenomics Database, Kyoto Encyclopedia of Genes and Genomes, Chemical Entitites of Biological Interest</td>
</tr>
<tr>
<td>Diseases</td>
<td>Medical Subject Headings, Systematized Nomenclature of Medicine</td>
</tr>
<tr>
<td>Literature</td>
<td>NCBI</td>
</tr>
<tr>
<td>Pathways</td>
<td>BioPax, Pathways Interaction Database, Reactome, Biocarta</td>
</tr>
<tr>
<td>Variants</td>
<td>dbSNP, Alfred, HuGENet, JSnp, HapMap, Seattle SNPs, Affymetrix</td>
</tr>
<tr>
<td>Structures</td>
<td>Protein Data Bank, NCBI</td>
</tr>
<tr>
<td>Other</td>
<td>Pharmacogenetics Research Network, US FDA, Centers for Disease Control and Prevention, GAPPNet</td>
</tr>
</tbody>
</table>
Medical Genetics Summaries
https://www.ncbi.nlm.nih.gov/books/NBK61999/

Medical Genetics Summaries is a growing collection of summaries which describe the impact that specific sequence variations have on health. The summaries review genetic variants that underlie inherited conditions, affect the risk of developing a disease in the future, or influence how an individual may respond to a specific drug.

Contents

Acknowledgments

Introduction
IUPHAR/BPS GUIDE TO PHARMACOLOGY database
http://www.guidetopharmacology.org/
IUPHAR/BPS GUIDE TO PHARMACOLOGY database

The IUPHAR/BPS Guide to PHARMACOLOGY provides expert-curated *molecular interactions between successful and potential drugs and their targets in the human genome*


Useful links: http://www.guidetopharmacology.org/links.jsp

SlideShare - Users will find presentations that include descriptions of content, mining approaches and utilities that extend beyond what is documented on the site. There is also added a set of generic slides which can be used by anyone presenting or teaching on GtoPdb.
https://www.slideshare.net/GuidetoPHARM
ORPHANET is a free database dedicated to information on rare diseases and orphan drugs

http://www.orpha.net/
DrugBank

DrugBank is a detailed database on small molecule and biotech drugs. Each drug entry ("DrugCard") includes extensive information on properties, structure, and biology (what the drug does in the body)

https://www.drugbank.ca/
Nucleic Acids Research. Database issue
https://academic.oup.com/nar

Database issue
The issue includes 152 papers, of which 54 describe newly created databases. 82 papers provide updates on databases that have been described in the previous NAR Database Issues, and 16 contain updates on databases whose descriptions have previously been published in other journals.

View the issue

2016 Web Server issue
The papers in this issue reflect current and emerging trends in bioinformatics and computational biology. Of particular interest are recent works which move beyond an understanding of bio-molecular phenomena toward an emphasis on the development of engineering tools to exploit that understanding.

View the issue

Email alerts
Register to receive advance article and new issue alerts as soon as new content of Nucleic Acids Research is published online.
Altre basi dati di interesse disciplinare (3)

- Evidence-based Drug Resources (Clinical trials, Systematic reviews, EBM)
Free Resources

PROSPERO

PROSPERO *international prospective register of systematic reviews* is an international database of prospectively registered systematic reviews in *health and social care, welfare, public health, education, crime, justice, and international development, where there is a health related outcome*. Reviews of reviews and reviews of methodological issues that contain at least one outcome of direct patient or clinical relevance are also included.

https://www.crd.york.ac.uk/PROSPERO/#index.php

ClinicalTrials.gov

The U.S. NIH database ClinicalTrials.gov ClinicalTrials.gov is a registry and results database of *publicly and privately supported clinical studies of human participants conducted around the world*. Registers trials that are recruiting and reports which have been completed. *A majority of the trials in this registry are never published*. Access from PubMed homepage too.

https://clinicaltrials.gov/
Free Resources

European Clinical Trials Database (EudraCT)
EudraCT V10, marks the final step of a process through which summary clinical trial results will be made publicly available through the EU Clinical Trials Register (EU CTR)
https://eudract.ema.europa.eu/

EUDRAPHARM
EudraPharm is intended to be a source of information on all medicinal products for human use that have been authorised in the European Union (EU) and the European Economic Area (EEA) and information on clinical trials of medicinal products including products with or without a marketing authorization
Free Resources

**ISRCTN Registry**

is a primary *clinical trial registry* recognised by **WHO** and **ICMJE** (International Committee of Medical Journal Editors) that accepts *all clinical research studies (whether proposed, ongoing or completed)*, providing content validation and curation and the unique identification number necessary for publication.

*All study records in the database are freely accessible and searchable.*

The **ISRCTN registry** is administered and the database is hosted and published on behalf of ISRCTN by BioMed Central, a publisher of open access peer-reviewed biomedical journals.

WHO ICTPR Search Portal

The Clinical Trials Search Portal provides access to a central database containing the trial registration data sets provided by the registries of: Australian New Zealand Clinical Trials Registry (ANZCTR); Brazilian Clinical Trials Registry (ReBec); Chinese Clinical Trial Register (ChiCTR); Clinical Research Information Service (CRiS), Republic of Korea; ClinicalTrials.gov; Clinical Trials Registry - India (CTRI); Cuban Public Registry of Clinical Trials (RPCEC); EU Clinical Trials Register (EU-CTR); German Clinical Trials Register (DRKS); Iranian Registry of Clinical Trials (IRCT); ISRCTN.org; Japan Primary Registries Network (JPRN); Pan African Clinical Trial Registry (PACTR); Peruvian Clinical Trials Registry (REPEC); Sri Lanka Clinical Trials Registry (SLCTR); Thai Clinical Trials Register (TCTR); The Netherlands National Trial Register (NTR).

It also provides links to the full original records.

http://apps.who.int/trialsearch/Default.aspx
Altre basi dati di interesse disciplinare (3)

- Legal Databases
Databases

Leggi d’Italia
http://studiolegale.leggiditalia.it/

IUS Explorer
IUS EXPLORER is a new Giuffrè search engine that allows an integrated access to all the online resources published by Giuffré
https://www.iusexplorer.it/

Gazzetta Ufficiale
The archives of the Official Journal allow you to search and browse in an integral and free way the different Series of this journal and it also allows to search and view the text of all state legislation published since 1946, constantly updated with the changes made daily in other legislation.
http://www.gazzettaufficiale.it/index.jsp
Free Databases

**EUR-Lex – Normativa UE**

EUR-Lex is the portal to European Union law. In addition to the current selection of texts on the homepage, there are three ways in which information can be accessed: consulting the Official Journal, searching by criteria and consulting the collections. There collections are: - the Official Journal, published every working day in 20 languages, consists of three series: L (all binding legislation), C (information, preparatory work, notices and recommendations) and a supplement S (tenders) published in the TED database; - a Search engine: a large selection of search criteria is available, search by word, by reference or by document number; - a number of Collections: Treaties, Legislation, Legislation in preparation, Case law, Parliamentary questions.

http://eur-lex.europa.eu/homepage.html

**The European Legal Database on Drugs (ELDD)**

The European Legal Database on Drugs (ELDD) is a free database maintained by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) with the assistance of a network of national 'legal correspondents', and provides information on European drugs-related legislation.

http://www.emcdda.europa.eu/eldd
How to find databases and bibliographic resources?

Biblioteca Digitale di Ateneo – Portale AIRE
http://metaricerca.cab.unipd.it:8332/V/

PER ALTRE FONTI INFORMATIVE DI INTERESSE!
References

Guides and Tutorials

NIH U.S. National Library of Medicine. Subject Guides on Selected Topics.  
https://www.nlm.nih.gov/services/Subject_Guides/subjectguidesonselectedtopics/  
NLM Guides, Tutorials and Quick Tours  
NCBI Learn  
NCBI Handbook  
https://www.ncbi.nlm.nih.gov/books/NBK143764/  
TOXNET Tutorial & Recordings  
https://nnlm.gov/nto/professional-development/tutorials-recordings/toxnet-tutorials-recordings  
NLM learning Resources Database  
https://learn.nlm.nih.gov/
References

Databases


https://doi.org/10.1093/nar/gkv1290


https://doi.org/10.1093/nar/gkv1037


https://doi.org/10.1093/nar/gkv951
References

Databases
http://www.futuremedicine.com/doi/full/10.2217/pgs.10.15

Consultare la rivista Nucleic Acids Research che pubblica annualmente un Database issue con aggiornamento dei free database esistenti e inserendo i nuovi database
https://academic.oup.com/nar
L’ultimo pubblicato nel 2016: https://academic.oup.com/nar/issue/45/D1

eBooks