**Databases**

BioProject (formerly Genome Project)

A collection of genomics, functional genomics, and genetics studies and links to their resulting datasets. This resource describes project scope, material, and objectives and provides a mechanism to retrieve datasets that are often difficult to find due to inconsistent annotation, multiple independent submissions, and the varied nature of diverse data types which are often stored in different databases.

Все геномные проекты, законченные и начатые ( **61358)**

**BioSample**

The BioSample database contains descriptions of biological source materials used in experimental assays.

**BioSystems**

Database that groups biomedical literature, small molecules, and sequence data in terms of biological relationships.

**Bookshelf**

A collection of biomedical books that can be searched directly or from linked data in other NCBI databases. The collection includes biomedical textbooks, other scientific titles, genetic resources such as GeneReviews, and NCBI help manuals.

**ClinVar**

A resource under development to provide a public, tracked record of reported relationships among human variation and observed health status. ClinVar has a projected launch date for late 2011.

**CloneDB (formerly Clone Registry)**

A database that integrates information about clones and libraries, including sequence data, map positions and distributor information.

**Computational Resources from NCBI's Structure Group**

A centralized page providing access and links to resources developed by the Structure Group of the NCBI Computational Biology Branch (CBB). These resources cover databases and tools to help in the study of macromolecular structures, conserved domains and protein classification, small molecules and their biological activity, and biological pathways and systems.

**Consensus CDS (CCDS)**

A collaborative effort to identify a core set of human and mouse protein coding regions that are consistently annotated and of high quality.

Гомологичные гены из геномов мыши и человека 18,535 генов

**Conserved Domain Database (CDD)**

A collection of sequence alignments and profiles representing protein domains conserved in molecular evolution. It also includes alignments of the domains to known 3-dimensional protein structures in the MMDB database.

**Database of Expressed Sequence Tags (dbEST)**

A divison of GenBank that contains short single-pass reads of cDNA (transcript) sequences. dbEST can be searched directly through the Nucleotide EST Database.

**Database of Genome Survey Sequences (dbGSS)**

A division of GenBank that contains short single-pass reads of genomic DNA. dbGSS can be searched directly through the Nucleotide GSS Database.

**Database of Genomic Structural Variation (dbVar)**

The dbVar database has been developed to archive information associated with large scale genomic variation, including large insertions, deletions, translocations and inversions. In addition to archiving variation discovery, dbVar also stores associations of defined variants with phenotype information.

Крупные вариации (>1 kb) в геноме и ассоциированные фенотипы для 10 эукариот

**Database of Genotypes and Phenotypes (dbGaP)**

An archive and distribution center for the description and results of studies which investigate the interaction of genotype and phenotype. These studies include genome-wide association (GWAS), medical resequencing, molecular diagnostic assays, as well as association between genotype and non-clinical traits.

**Database of Major Histocompatibility Complex (dbMHC)**

An open, publicly accessible platform where the HLA community can submit, edit, view, and exchange data related to the human major histocompatibility complex. It consists of an interactive Alignment Viewer for HLA and related genes, an MHC microsatellite database, a sequence interpretation site for Sequencing Based Typing (SBT), and a Primer/Probe database.

**Database of Short Genetic Variations (dbSNP)**

Includes single nucleotide variations, microsatellites, and small-scale insertions and deletions. dbSNP contains population-specific frequency and genotype data, experimental conditions, molecular context, and mapping information for both neutral variations and clinical mutations.

Локальные мутации в геноме и ассоциированные заболевания/особенности организма **63,222,718**

**Epigenomics**

This resource enables users to explore and visualize richly-annotated epigenomics datasets. It provides a unique interface to search and navigate epigenomic data in the context of biological sample information, as well as tools to select, download and view multiple sets of epigenomic data as tracks on genome browsers.

Метилирование ДНК и модификации гистонов 3545 экспериментов

**GenBank**

The NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis. GenBank consists of several divisions, most of which can be accessed through the Nucleotide database. The exceptions are the EST and GSS divisions, which are accessed through the Nucleotide EST and Nucleotide GSS databases, respectively.

Посл ДНК/РНК

**Gene**

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 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.

Инфо о генах

**Gene Expression Omnibus (GEO) Database**

A public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted and tools are provided to help users query and download experiments and curated gene expression profiles.

Экспрессионные чипы

Gene Expression Omnibus (GEO) Datasets

Stores curated gene expression and molecular abundance DataSets assembled from the Gene Expression Omnibus (GEO) repository. DataSet records contain additional resources, including cluster tools and differential expression queries.

**Gene Expression Omnibus (GEO) Profiles**

Stores individual gene expression and molecular abundance Profiles assembled from the Gene Expression Omnibus (GEO) repository. Search for specific profiles of interest based on gene annotation or pre-computed profile characteristics.

**GeneReviews**

A collection of expert-authored, peer-reviewed disease descriptions on the NCBI Bookshelf that apply genetic testing to the diagnosis, management, and genetic counseling of patients and families with specific inherited conditions.

**GeneTests**

A publicly funded medical genetics information resource developed for physicians, other healthcare providers, and researchers, available at no cost to all interested persons.

**Genes and Disease**

Summaries of information for selected genetic disorders with discussions of the underlying mutation(s) and clinical features, as well as links to related databases and organizations.

**Genetic Testing Registry (GTR)**

A voluntary registry of genetic tests and laboratories, with detailed information about the tests such as what is measured and analytic and clinical validity. GTR also is a nexus for information about genetic conditions and provides context-specific links to a variety of resources, including practice guidelines, published literature, and genetic data/information. The initial scope of GTR includes single gene tests for Mendelian disorders, as well as arrays, panels and pharmacogenetic tests.

**Genome**

Contains sequence and map data from the whole genomes of over 1000 organisms. The genomes represent both completely sequenced organisms and those for which sequencing is in progress. All three main domains of life (bacteria, archaea, and eukaryota) are represented, as well as many viruses, phages, viroids, plasmids, and organelles.

Genome Information by organism. Overview [8667]Eukaryotes [2226]Prokaryotes [14895]Viruses [3343]
В т.ч. незаконченные.

**Genome Reference Consortium (GRC)**

The Genome Reference Consortium (GRC) maintains responsibility for the human and mouse reference genomes. Members consist of The Genome Center at Washington University, the Wellcome Trust Sanger Institute, the European Bioinformatics Institute (EBI) and the National Center for Biotechnology Information (NCBI). The GRC works to correct misrepresented loci and to close remaining assembly gaps. In addition, the GRC seeks to provide alternate assemblies for complex or structurally variant genomic loci. At the GRC website (http://www.genomereference.org), the public can view genomic regions currently under review, report genome-related problems and contact the GRC.

**HIV-1, Human Protein Interaction Database**

A database of known interactions of HIV-1 proteins with proteins from human hosts. It provides annotated bibliographies of published reports of protein interactions, with links to the corresponding PubMed records and sequence data.

**HomoloGene**

A gene homology tool that compares nucleotide sequences between pairs of organisms in order to identify putative orthologs. Curated orthologs are incorporated from a variety of sources via the Gene database.

**Influenza Virus**

A compilation of data from the NIAID Influenza Genome Sequencing Project and GenBank. It provides tools for flu sequence analysis, annotation and submission to GenBank. This resource also has links to other flu sequence resources, and publications and general information about flu viruses.

Journals in NCBI Databases

Subset of the NLM Catalog database providing information on journals that are referenced in NCBI database records, including PubMed abstracts. This subset can be searched using the journal title, MEDLINE or ISO abbreviation, ISSN, or the NLM Catalog ID.

**MeSH Database**

MeSH (Medical Subject Headings) is the U.S. National Library of Medicine's controlled vocabulary for indexing articles for MEDLINE/PubMed. MeSH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts.

**NCBI C++ Toolkit Manual**

A comprehensive manual on the NCBI C++ toolkit, including its design and development framework, a C++ library reference, software examples and demos, FAQs and release notes. The manual is searchable online and can be downloaded as a series of PDF documents.

**NCBI Education Page**

Provides links to tutorials and training materials, including PowerPoint slides and print handouts.

**NCBI Glossary**

Part of the NCBI Handbook, this glossary contains descriptions of NCBI tools and acronyms, bioinformatics terms and data representation formats.

**NCBI Handbook**

An extensive collection of articles about NCBI databases and software. Designed for a novice user, each article presents a general overview of the resource and its design, along with tips for searching and using available analysis tools. All articles can be searched online and downloaded in PDF format; the handbook can be accessed through the NCBI Bookshelf.

**NCBI Help Manual**

Accessed through the NCBI Bookshelf, the Help Manual contains documentation for many NCBI resources, including PubMed, PubMed Central, the Entrez system, Gene, SNP and LinkOut. All chapters can be downloaded in PDF format.

**NCBI Website Search**

A database of static NCBI web pages, documentation, and online tools. These pages include such content as specialized online sequence analysis tools, back issues of newsletters, legacy resource description pages, sample code, and other miscellaneous resources. Searching this database is equivalent to a site search tool for the whole NCBI web site, with the exception of the FTP directories.

**National Library of Medicine (NLM) Catalog**

Bibliographic data for all the journals, books, audiovisuals, computer software, electronic resources and other materials that are in the library's holdings.

**Nucleotide Database**

A collection of nucleotide sequences from several sources, including GenBank, RefSeq, the Third Party Annotation (TPA) database, and PDB. Searching the Nucleotide Database will yield available results from each of its component databases.

**Online Mendelian Inheritance in Animals (OMIA)**

A database of genes, inherited disorders and traits in animal species (other than human and mouse), with textual information and references, as well as links to relevant records from other NCBI databases, such as PubMed and Gene.

**Online Mendelian Inheritance in Man (OMIM)**

A database of human genes and genetic disorders. NCBI maintains current content and continues to support its searching and integration with other NCBI databases. However, OMIM now has a new home at omim.org, and users are directed to this site for full record displays.

Наследственные заболевания человека

**PopSet**

Database of related DNA sequences that originate from comparative studies: phylogenetic, population, environmental and, to a lesser degree, mutational. Each record in the database is a set of DNA sequences. For example, a population set provides information on genetic variation within an organism, while a phylogenetic set may contain sequences, and their alignment, of a single gene obtained from several related organisms.

**Probe**

A public registry of nucleic acid reagents designed for use in a wide variety of biomedical research applications, together with information on reagent distributors, probe effectiveness, and computed sequence similarities.

**Protein Clusters**

A collection of related protein sequences (clusters), consisting of Reference Sequence proteins encoded by complete prokaryotic and organelle plasmids and genomes. The database provides easy access to annotation information, publications, domains, structures, external links, and analysis tools.

**Protein Database**

A database that includes protein sequence records from a variety of sources, including GenPept, RefSeq, Swiss-Prot, PIR, PRF, and PDB.

**PubChem BioAssay**

Consists of deposited bioactivity data and descriptions of bioactivity assays used to screen the chemical substances contained in the PubChem Substance database, including descriptions of the conditions and the readouts (bioactivity levels) specific to the screening procedure.

**PubChem Compound**

Contains unique, validated chemical structures (small molecules) that can be searched using names, synonyms or keywords. The compound records may link to more than one PubChem Substance record if different depositors supplied the same structure. These Compound records reflect validated chemical depiction information provided to describe substances in PubChem Substance. Structures stored within PubChem Compounds 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.

**PubChem Substance**

PubChem Substance records contain substance information electronically submitted to PubChem by depositors. This includes any chemical structure information submitted, as well as chemical names, comments, and links to the depositor's web site.

**PubMed**

A database of citations and abstracts for biomedical literature from MEDLINE and additional life science journals. Links are provided when full text versions of the articles are available via PubMed Central (described below) or other websites.

**PubMed Central (PMC)**

A digital archive of full-text biomedical and life sciences journal literature, including clinical medicine and public health.

**PubMed Health**

A collection of clinical effectiveness reviews and other resources to help consumers and clinicians use and understand clinical research results. These are drawn from the NCBI Bookshelf and PubMed, including published systematic reviews from organizations such as the Agency for Health Care Research and Quality, The Cochrane Collaboration, and others (see complete listing). Links to full text articles are provided when available.

**RefSeqGene**

A collection of human gene-specific reference genomic sequences. RefSeq gene is a subset of NCBI's RefSeq database, and are defined based on review from curators of locus-specific databases and the genetic testing community. They form a stable foundation for reporting mutations, for establishing consistent intron and exon numbering conventions, and for defining the coordinates of other biologically significant variation. RefSeqGene is a part of the Locus Reference Genomic (LRG) Collaboration.

**Reference Sequence (RefSeq)**

A collection of curated, non-redundant genomic DNA, transcript (RNA), and protein sequences produced by NCBI. RefSeqs provide a stable reference for genome annotation, gene identification and characterization, mutation and polymorphism analysis, expression studies, and comparative analyses. The RefSeq collection is accessed through the Nucleotide and Protein databases.

**Retrovirus Resources**

A collection of resources specifically designed to support the research of retroviruses, including a genotyping tool that uses the BLAST algorithm to identify the genotype of a query sequence; an alignment tool for global alignment of multiple sequences; an HIV-1 automatic sequence annotation tool; and annotated maps of numerous retroviruses viewable in GenBank, FASTA, and graphic formats, with links to associated sequence records.

**SARS CoV**

A summary of data for the SARS coronavirus (CoV), including links to the most recent sequence data and publications, links to other SARS related resources, and a pre-computed alignment of genome sequences from various isolates.

**Sequence Read Archive (SRA)**

The Sequence Read Archive (SRA) stores sequencing data from the next generation of sequencing platforms including Roche 454 GS SystemR, Illumina Genome AnalyzerR, Life Technologies AB SOLiD SystemR, Helicos Biosciences HeliscopeR, Complete GenomicsR, and Pacific Biosciences SMRTR.

**Structure (Molecular Modeling Database)**

Contains macromolecular 3D structures derived from the Protein Data Bank, as well as tools for their visualization and comparative analysis.

**Taxonomy**

Contains the names and phylogenetic lineages of more than 160,000 organisms that have molecular data in the NCBI databases. New taxa are added to the Taxonomy database as data are deposited for them.

**Third Party Annotation (TPA) Database**

A database that contains sequences built from the existing primary sequence data in GenBank. The sequences and corresponding annotations are experimentally supported and have been published in a peer-reviewed scientific journal. TPA records are retrieved through the Nucleotide Database.

**Trace Archive**

A repository of DNA sequence chromatograms (traces), base calls, and quality estimates for single-pass reads from various large-scale sequencing projects.

**UniGene**

A database that provides sets of transcript sequences that appear to come from the same transcription locus (gene or expressed pseudogene), together with information on protein similarities, gene expression, cDNA clone reagents, and genomic location.

**UniGene Library Browser**

This database contains libraries of Expressed Sequence Tags (ESTs) organized by organism, tissue type and developmental stage.

**UniSTS**

A comprehensive database of sequence tagged sites (STSs) derived from STS-based maps and other experiments. STSs are defined by PCR primer pairs and are associated with additional information, such as genomic position, genes, and sequences.

**Viral Genomes**

A wide range of resources, including a brief summary of the biology of viruses, links to viral genome sequences in Entrez Genome, and information about viral Reference Sequences, a collection of reference sequences for thousands of viral genomes.

**Virus Variation**

An extension of the Influenza Virus Resource to other organisms, providing an interface to download sequence sets of selected viruses, analysis tools, including virus-specific BLAST pages, and genome annotation pipelines.