(mostly from Rocha and Ferreira Bioinformatics algorithms, 2018)

Sequence data

The European Molecular Biology Laboratory (EMBL)-bank was created in 1982 as the first international database for nucleotide sequences. Also, in that year, there was the public release of Genbank (www.ncbi.nlm.nih.gov/genbank/), now maintained by the National Center for Biotechnology Information (NCBI) in the United States, which contained an annotated collection of public available nucleotide sequences and respective sequence translations.

In 1986, the Swiss-Prot (now part of the UniProtKB) database was presented, containing nonredundant and curated protein sequence data complemented with other high level information and interconnected with other sequence resources.

In 1988, the International Nucleotide Sequence Database Collaboration (INSDC) was launched, a joint effort of EMBL-EBI in Europe, NCBI in the United States and DDBJ (www.ddbj.nig.ac.jp/) in Japan to collect and disseminate nucleotide sequences. It currently involves the databases of DNA Data Bank of Japan, GenBank and the European Nucleotide Archive (ENA, www.ebi.ac.uk/ena).

ENA (includes EMBL-bank) – www.ebi.ac.uk/ena
GenBank – www.ncbi.nlm.nih.gov/GenBank
www.ncbi.nlm.nih.gov/refseq
DDBJ – www.ddbj.nig.ac.jp
NCBI Gene – http://www.ncbi.nlm.nih.gov/gene
NCBI RefSeq – http://www.ncbi.nlm.nih.gov/refseq
UniProtKB/Swiss-Prot (www.uniprot.org/uniprot/).
Gencode – www.gencodegenes.org
The Gencode annotation started as an effort within the ENCODE project to provide a fully integrated annotation of the human genome.

HUMAN REFERENCE GENOME https://www.ncbi.nlm.nih.gov/grc/human

dbSNP – www.ncbi.nlm.nih.gov/snp
dbVar – www.ncbi.nlm.nih.gov/dbvar

These are two databases from NCBI that contain the annotation of short genetic and large structural variations within the human genome and other species. dbSNP is mostly focused on point mutations, microsatellites, and small insertions and deletions. It contains information on the mutated alleles, their sequence context, visualization of their occurrence within the gene sequence, frequency in populations and also connects with other databases to show information on clinical significance. dbVar entries contain a view and details of the genomic region where the variation occurs, complemented experimental evidence and validation, publication where its was first reported and clinical associations.

ClinVar – www.ncbi.nlm.nih.gov/clinvar/
ClinVar is a database that provides information and supporting evidence on the association of human genetic variation and phenotypes. It is particularly useful in the clinical and health context since it reports variants found in patient samples along with assertions made by the researchers or the clinicians that submitted data about the clinical relevance of these variants.

Browsers

The Ensembl genome database (www.ensembl.org/), a joint initiative of the European Bioinformatics Institute and the Wellcome Trust Sanger Institute, and the genome browser at University of California Santa Cruz (UCSC, genome.ucsc.edu/) are the most well known browsers for genomic information. These browsers have evolved to very complete resources, integrating nowadays many different types of genomic information from several different species.

UCSC Genome Browser – https://genome.ucsc.edu/
Ensembl – http://www.ensembl.org/

Structures

Protein Data Bank (PDB) – http://www.rcsb.org/
PDB is a database that contains structural data of proteins, nucleic acids and other complex assemblies. It provides functionalities for data deposit and download and tools for data visualizations in multiple data formats.
It contains information organized by protein, containing from the sequence, annotations of secondary structure, tri-dimensional coordinates and views, similarity search at sequence and structure level and the details of the experiment

Gene Expression

Gene Expression Ominbus (GEO) – www.ncbi.nlm.nih.gov/geo/
GEO is a database from NCBI that collects gene expression datasets obtained either with micro-array or sequencing
technologies.

Literature

PubMed – www.ncbi.nlm.nih.gov/pubmed/
PubMed is a database from NCBI that indexes information of scientific articles related to biomedical and life sciences research. Article entries contain links to publisher website. Advanced article search can be made based on keywords, title or author names.