



Human Genome Resources

National Center for Biotechnology Information ■ National Library of Medicine ■ 8600 Rockville Pike ■ Bethesda, MD 20894

NCBI has an ongoing program to assemble and annotate the human genome, incorporating updates as new and revised genome data are deposited in public resources. The data are available to the public through a number of NCBI services. The Human Genome page (www.ncbi.nlm.nih.gov/genome/guide/human) serves as an online guide to the human genome resources. Key services are highlighted here.

Genes

Entrez Gene—curated central hub of information about genes and loci. Stable ID for each locus, and links to a diverse array of resources, including nomenclature, reference sequences, literature, mapping information, related databases. A collaborative effort among NCBI, Human Gene Nomenclature Committee, OMIM, and others. Also supports a joint effort to track the identification of novel genes.

www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene

OMIM—Online Mendelian Inheritance in Man—continuously updated catalog of human genes and genetic disorders,

with links to associated literature references, sequence records, maps, and related databases; one of the Entrez databases. www.ncbi.nlm.nih.gov/Entrez

RefSeq—see Sequences

UniGene—see Sequences

HomoloGene—tool for comparing nucleotide sequences between pairs of organisms to identify putative orthologs. www.ncbi.nlm.nih.gov/HomoloGene

Sequences

dbEST database of Expressed Sequence Tags—short (about 300-500 bp) cDNA sequences representing single-pass reads from mRNA. Usually produced in large numbers and represent a snapshot of the genes expressed in a given tissue, and/or at a given developmental stage. Also includes ESTs generated by the Cancer Genome Anatomy Project (CGAP) (see **Cancer Research**). www.ncbi.nlm.nih.gov/dbEST

Entrez—provides integrated access to nucleotide and protein sequence data in GenBank, EMBL, DDBJ, RefSeq, PIR-International, PRF, Swiss-Prot, and PDB, along with 3D protein structures, gene mapping and phenotype information, and PubMed. Contains pre-computed similarity searches for each database record, producing a list of related sequences, structures, and literature records. Includes sequence data from >204,000 species; use the organism field to limit searches to human records. www.ncbi.nlm.nih.gov/Entrez

Human Genome Sequencing—sequencing progress of the Human Genome Project, international sequencing center data, links to constructed genomic contigs in Map Viewer, BLAST against chromosomes www.ncbi.nlm.nih.gov/genome

RefSeq—curated, non-redundant set of reference sequences including constructed genomic DNA contigs, mRNAs and proteins for known genes, and, in the future, entire chromosomes. Reference sequences are available through Entrez Gene, Entrez, BLAST, and the NCBI ftp site. www.ncbi.nlm.nih.gov/RefSeq

UniGene—ESTs and full-length mRNA sequences organized into clusters representing unique known or putative human genes. Annotated with mapping and expression information and cross-references to other resources. www.ncbi.nlm.nih.gov/UniGene

Genome Maps

Entrez—the Genomes domain of Entrez graphically represents genetic, cytogenetic, physical, and sequence maps that have been integrated to show common markers. Each human chromosome can be viewed in its entirety or explored in progressively greater detail. www.ncbi.nlm.nih.gov/Entrez

OMIM Gene Map—cytogenetic locations of genes reported in the literature and determined by a variety of mapping methods. Can be searched by gene symbol or cytogenetic chromosomal location. Accessible from the OMIM page (see **Genes**).

Human-Mouse Maps—comparative Map View displays of genes in homologous segments of DNA from human, mouse, and rat. www.ncbi.nlm.nih.gov/Homology

Human Genome Map Viewer—provides integrated views of chromosome maps. Displays maps that have been aligned to each other based on shared marker and gene names, or a common sequence coordinate system. Includes >20 sequence, cytogenetic, genetic linkage, radiation hybrid, SNP, and other maps, 9 of which can be viewed simultaneously. Some sequence maps are based on the contigs built from the draft and finished sequence data generated by the Human Genome Project. www.ncbi.nlm.nih.gov/mapview

OMIM Morbid Map—see Disorders.

Mitelman Chromosomal Aberration Summary—see Cytogenetics

Mapped Markers

UniSTS—short (about 200-500 bp) genomic sequences that are thought to be operationally unique in a genome, and therefore define a specific position on the physical map. www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists

E-PCR Electronic PCR—find putative map location of a query

sequence. Computational procedure for finding sequence tagged sites in DNA sequences. www.ncbi.nlm.nih.gov/sutils/e-pcr

Human Genome MapViewer—see Genome Maps

OMIM Gene Map—see Genome Maps

Genetic Variations

dbSNP—Database of Single Nucleotide Polymorphisms—database of single nucleotide polymorphisms, microsatellites, and small-scale insertions and deletions. Contains population-specific frequency and genotype data, experimental conditions, molecular context, and mapping information for both neutral polymorphisms and clinical mutations. www.ncbi.nlm.nih.gov/projects/SNP

OMIM—allelic variants in ~900 (9%) of OMIM records. To view a list of those OMIM records, check “Only Records with: Allelic Variants” on the Limits page (see Genes).

Entrez Gene Mutation Databases—links to numerous mutation databases from the OMIM home page and from related Genes entries (see Genes).

Cytogenetics

Human BAC Resource—Cytogenetic resource of large-insert, FISH-mapped, clones containing sequence-tagged sites. Will help integrate cytogenetic, radiation-hybrid, linkage, and sequence maps of the human genome. Includes links to clone distributors. www.ncbi.nlm.nih.gov/genome/cyto/hbrc.shtml

Mitelman Chromosomal Aberration—Map View display of chromosomal breakpoints in human cancer, by Drs. Mitelman, Mertens, and Johansson. Associated with CGAP (see Gene Expression).

OMIM Morbid Map—see Disorders

SKY/CGH—Spectral Karyotyping and Comparative Genomic Hybridization Database—repository of publicly submitted data from SKY and CGH, complementary fluorescent molecular cytogenetic techniques. SKY facilitates identification of chromosomal aberrations; CGH can be used to generate a map of DNA copy number changes in tumor genomes. www.ncbi.nlm.nih.gov/sky

Gene Expression

CGAP—Cancer Genome Anatomy Project—interdisciplinary program to identify the human genes expressed in different cancerous states, based on cDNA (EST) libraries, and to determine the molecular profiles of normal, precancerous, and malignant cells. Collaboration among the National Cancer Institute, the NCBI, and numerous research labs. www.ncbi.nlm.nih.gov/projects/CGAP

CGAP cDNA Expression Profiles—distribution of ESTs by UniGene cluster and cDNA (EST) library. Accessible from the CGAP page. www.ncbi.nlm.nih.gov/projects/CGAP

SAGEmap—Differential analysis of CGAP SAGE (Serial Analysis of Gene Expression)—libraries. Also, comprehensive analysis of SAGE tags in human GenBank records; UniGene identifier is assigned to each human sequence that contains a SAGE tag. www.ncbi.nlm.nih.gov/projects/SAGE

GEO—Gene Expression Omnibus—repository for gene expression data from any organism or artificial source. Includes data from platforms such as spotted microarray, high-density oligonucleotide array, hybridization filter, and serial analysis of gene expression. www.ncbi.nlm.nih.gov/projects/geo

Disorders

Genes and Diseases—introduction to the relationship between genetic factors and human disease. Summary information for ~80 genetic diseases with links to related databases and organizations. www.ncbi.nlm.nih.gov/disease

Mitelman Chromosomal Aberration Summary—see Cytogenetics

OMIM—see Genes

OMIM Morbid Map—alphabetical listing of diseases and corresponding cytogenetic map locations, with links to OMIM entries. Accessible from OMIM page (see Genes).

