

## Global Core Biodata Resources

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2023
<a href="#">Alliance of Genome Resources</a>	USA, Canada, UK	The primary mission of the Alliance of Genome Resources (the Alliance) is to develop and maintain sustainable genome information resources that facilitate the use of diverse model organisms in understanding the genetic and genomic basis of human biology, health and disease.	US National Institutes of Health (NIH)
<a href="#">BacDive</a>	Germany	BacDive is the worldwide largest database for standardized bacterial information. Its mission is to mobilize and integrate research data on strain level from diverse sources and make it freely accessible.	German Federal Ministry of Education and Research (BMBF)
<a href="#">BAR: Bio-Analytic Resource for Plant Biology</a>	Canada	The Bio-Analytic Resource for Plant Biology (BAR) offers a combination of data visualization tools and database web services for exploring a wide variety of large data sets covering multiple levels of plant biology.	Genome Canada, National Science and Engineering Research Council (Canada), University of Toronto
<a href="#">Bgee</a>	Switzerland	Bgee is a database for retrieval and comparison of gene expression patterns across multiple animal species. It provides an intuitive answer to the question "where is a gene expressed?" and supports research in cancer and agriculture as well as evolutionary biology.	Swiss State Secretariat for Education, Research, and Innovation (SERI), European Commission, Swiss National Science Foundation (SNSF), SwissUniversities (CHORD)
<a href="#">BRENDA</a>	Germany	BRENDA is the main collection of enzyme functional data available to the scientific community.	German Federal Ministry of Education and Research (BMBF)
<a href="#">Catalogue of Life</a>	Netherlands, China, Denmark, UK, USA	Catalogue of Life (COL) is a collaboration and a data resource bringing together the effort and contributions of taxonomists and informaticians from around the world. COL aims to address the needs of researchers, policy-makers, environmental managers and the wider public for a consistent and up-to-date listing of all the world's known species. COL also supports those who need to manage their own taxonomic information and species lists.	European Commission, Naturalis Biodiversity Center, University of Illinois, Chinese Academy of Sciences, Global Biodiversity Information Facility, Ministry of Education, Culture, and Science, Netherlands, Smithsonian Institution, Species 2000, Stichting ter Bevordering van Natuurwetenschappelijk Onderzoek
<a href="#">CATH</a>	UK	The CATH database is a free, publicly available online resource that provides information on the evolutionary relationships of protein domains. The domains are classified within the CATH structural hierarchy: at the Class (C), Architecture (A), Topology/fold (T), and Homologous superfamily (H) levels.	UK Research and Innovation (UKRI), Wellcome

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2023
<a href="#">Cellosaurus</a>	Switzerland	The Cellosaurus is a knowledge resource on cell lines. It aims to describe all cell lines used in biomedical research. Its scope encompasses both vertebrates and invertebrates.	Swiss State Secretariat for Education, Research, and Innovation (SERI)
<a href="#">ChEBI</a>	UK	Chemical Entities of Biological Interest (ChEBI) is a freely available dictionary of molecular entities focused on 'small' chemical compounds.	European Commission, EMBL, UK Research and Innovation (UKRI)
<a href="#">ChEMBL</a>	UK	ChEMBL is a manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs.	Wellcome, EMBL, UK Research and Innovation (UKRI)
<a href="#">CIViC: Clinical Interpretation of Variants in Cancer</a>	USA	A community-edited forum for discussion and interpretation of peer-reviewed publications pertaining to the clinical relevance of variants (or biomarker alterations) in cancer.	US National Institutes of Health (NIH)
<a href="#">Clinical Genome Resource</a>	USA	ClinGen is a National Institutes of Health (NIH)-funded resource dedicated to building a central resource that defines the clinical relevance of genes and variants for use in precision medicine and research.	US National Institutes of Health (NIH)
<a href="#">DNA Data Bank of Japan</a>	Japan	DDBJ Center collects nucleotide sequence data as a member of INSDC (International Nucleotide Sequence Database Collaboration) and provides freely available nucleotide sequence data and supercomputer system, to support research activities in life science.	Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Science and Technology Agency (JST)
<a href="#">EcoCyc</a>	USA	EcoCyc is a scientific database for the bacterium Escherichia coli K-12 MG1655. The EcoCyc project performs literature-based curation of its genome, and of transcriptional regulation, transporters, and metabolic pathways.	US National Institutes of Health (NIH)
<a href="#">Ensembl</a>	UK	Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data.	Wellcome, UK Research and Innovation (UKRI), European Commission, EMBL, US National Institutes of Health (NIH), Save the Tasmanian Devil Program Appeal
<a href="#">Europe PMC</a>	UK	Europe PMC provides comprehensive access to life sciences literature from trusted sources. It's available to anyone, anywhere for free.	Europe PMC Funders Group (see: <a href="https://europepmc.org/Funders/">https://europepmc.org/Funders/</a> )

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2023
<a href="#">European Nucleotide Archive</a>	UK	The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation.	EMBL, Chan Zuckerberg Initiative, European Commission, Gordon and Betty Moore Foundation, UK Research and Innovation (UKRI), Wellcome
<a href="#">FlyBase</a>	USA, UK	FlyBase is a database of genetic and molecular data for <i>D. melanogaster</i> and other <i>Drosophila</i> species	US National Institutes of Health (NIH), US National Science Foundation (NSF), Wellcome, UK Research and Innovation (UKRI)
<a href="#">GENCODE</a>	UK, Germany, Japan, Singapore, USA	The goal of the GENCODE project is to identify and classify all gene features in the human and mouse genomes with high accuracy based on biological evidence, and to release these annotations for the benefit of biomedical research and genome interpretation.	US National Institutes of Health (NIH), UK Research and Innovation (UKRI), British Council, EMBL
<a href="#">Gene Ontology</a>	USA, UK	The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research.	US National Institutes of Health (NIH)
<a href="#">Global Biodiversity Information Facility</a>	Denmark	GBIF—the Global Biodiversity Information Facility—is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.	GBIF Participant Network (see <a href="https://www.gbif.org/the-gbif-network">https://www.gbif.org/the-gbif-network</a> ), Danish Government, US National Science Foundation, European Commission, Bill & Melinda Gates Foundation
<a href="#">gnomAD</a>	USA	The Genome Aggregation Database (gnomAD) is a resource developed by an international coalition of investigators, with the goal of aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community.	US National Institutes of Health (NIH), Wellcome, Michael J Fox Foundation
<a href="#">GSA: Genome Sequence Archive</a>	China	The Genome Sequence Archive (GSA) is a data repository for collecting, archiving, managing and sharing raw genome sequence data.	Ministry of Science and Technology of the People's Republic of China, Ministry of Finance of the People's Republic of China, Chinese Academy of Sciences, International Union of Biological Sciences, Alliance of International Science Organizations, National Natural Science Foundation of China

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2023
<a href="#">GWAS Catalog</a>	UK, USA	The NHGRI-EBI Catalog of human genome-wide association studies (GWAS) provides a consistent, searchable, visualisable and freely available database of SNP-trait associations, which can be easily integrated with other resources, and is accessed by scientists, clinicians and other users worldwide.	EMBL, US National Institutes of Health (NIH)
<a href="#">GXD</a>	USA	The Gene Expression Database (GXD) is a community resource for gene expression information from the laboratory mouse. GXD stores and integrates different types of expression data and makes these data freely available in formats appropriate for comprehensive analysis. There is particular emphasis on endogenous gene expression during mouse development.	US National Institutes of Health (NIH)
<a href="#">HGNC: HUGO Gene Nomenclature Committee</a>	UK	The HGNC is responsible for approving unique symbols and names for human loci, including protein coding genes, ncRNA genes and pseudogenes, to allow unambiguous scientific communication.	Wellcome, US National Institutes of Health (NIH)
<a href="#">Human Disease Ontology Knowledgebase</a>	USA	The Disease Ontology Knowledgebase (DO-KB) provides access for exploring integrated disease mechanism and feature knowledge, semantically-defined within the Human Disease Ontology, serving as a reference framework for multiscale biomedical data integration and analysis within a unifying etiology-based disease classification.	US National Institutes of Health (NIH)
<a href="#">Human Protein Atlas</a>	Sweden	The Human Protein Atlas aims to map all the human proteins in cells, tissues, and organs using an integration of various omics technologies, including antibody-based imaging, mass spectrometry-based proteomics, transcriptomics, and systems biology. All the data in the knowledge resource is open access to allow scientists both in academia and industry to freely access the data for exploration of the human proteome.	Knut and Alice Wallenberg Foundation (SE) Swedish Research Council Erling Persson Foundation (SE)

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2023
<a href="#">IMEx: International Molecular Exchange Consortium</a>	UK, Canada, France, Italy, USA	The IMEx Consortium provides data on physical interactions between biomolecules (proteins, nucleic acids, non-genome encoded bioentities such as small molecules, carbohydrates and protein complexes) across the entire taxonomic spectrum.	EMBL, Wellcome, US National Institutes of Health (NIH), Ontario Research Fund (CA), Natural Sciences Research Council (CA), Canada Foundation for Innovation, Swiss National Science Foundation (SNSF), Fondation pour la Recherche Médicale (FR), Institut Français de Bioinformatique (FR), Genome British Columbia (CA), Foundation for the National Institutes of Health (USA), Canadian Institutes of Health Research, The Italian Association for Cancer Research, Fondazione Human Technopole (IT)
<a href="#">InterPro</a>	UK	InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. To classify proteins in this way, InterPro uses predictive models, known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro consortium.	Wellcome, UK Research and Innovation (UKRI), EMBL, US National Institutes of Health (NIH)
<a href="#">IUPHAR/BPS Guide to PHARMACOLOGY</a>	UK	The BPS/IUPHAR Guide to PHARMACOLOGY is an expert-curated resource of ligand-activity-target relationships, the majority of which come from high-quality pharmacological and medicinal chemistry literature. It is intended as a “one-stop shop” portal to pharmacological information and its main aim is to provide a searchable database with quantitative information on drug targets and the prescription medicines and experimental drugs that act on them.	Antibiotic Research UK, British Pharmacological Society, Global Antibiotic Research and Development Partnership Italian Pharmacological Society, International Union of Basic and Clinical Pharmacology (IUPHAR), LifeArc (UK), University of Edinburgh, Medicines for Malaria Venture (CH)
<a href="#">LIPID MAPS</a>	UK, USA	LIPID Metabolites And Pathways Strategy (LIPID MAPS®) is a multi-institutional supported website and database that provides access to a large number of globally used lipidomics resources. LIPID MAPS® has internationally led the field of lipid curation, classification, and nomenclature since 2003. LIPID MAPS® generates tools and resources that address the emerging area of Systems Lipidomics, applied to biomedical research. This ranges from underpinning mechanistic studies on tissues, model systems and pathogenic and commensal microbes, through to human cohort epidemiology.	Wellcome, NIH, UKRI (MRC), Cayman Chemical, Avanti Polar Lipids, Merck Chemicals Ltd.

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2023
<a href="#">LPSN: List of Prokaryotic names with Standing in Nomenclature</a>	Germany	The List of Prokaryotic names with Standing in Nomenclature (LPSN) provides comprehensive information on the nomenclature of prokaryotes and related data.	Leibniz Institute DSMZ (DE), International Committee on Systematics of Prokaryotes (ICSP)
<a href="#">MGD: Mouse Genome Database</a>	USA	MGI is the international database resource for the laboratory mouse, providing integrated genetic, genomic, and biological data to facilitate the study of human health and disease.	US National Institutes of Health (NIH)
<a href="#">Orphadata Science</a>	France	Orphadata provides the scientific community with comprehensive, quality data sets related to rare diseases and orphan drugs from the Orphanet knowledge base, in reusable formats.	INSERM, European Commission, French Ministry of Health
<a href="#">PANTHER</a>	USA	The mission of the PANTHER knowledgebase is to support biomedical and other research by providing comprehensive information about the evolution of protein-coding gene families, particularly protein phylogeny, function and genetic variation impacting that function.	US National Science Foundation (NSF), UK Research and Innovation (UKRI), US National Institutes of Health (NIH)
<a href="#">PharmGKB: Pharmacogenomics Knowledgebase</a>	USA	PharmGKB is a comprehensive resource that curates knowledge about the impact of genetic variation on drug response for clinicians and researchers.	US National Institutes of Health (NIH)
<a href="#">Planteome</a>	USA	The Planteome Knowledgebase and Ontologies for Plant Biology provides a centralized web portal which features a suite of interconnected reference ontologies utilized by the plant biology community for the annotation of plant gene expression data, traits, phenotypes, genomes and germplasm. The Planteome database provides access to reference and species-specific ontologies and bioentities, or data objects, including proteins, genes, RNA transcripts, gene models and QTLs.	US National Science Foundation (NSF), United States Department of Agriculture (USDA), Oregon State University, CGIAR
<a href="#">PomBase</a>	UK	PomBase is a comprehensive database for the fission yeast <i>Schizosaccharomyces pombe</i> , providing structural and functional annotation, literature curation and access to large-scale data sets.	Wellcome

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2023
<a href="#">Protein Data Bank</a>	USA, Japan, UK	<p>Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.</p> <p>The Worldwide PDB (wwPDB) organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.</p>	Japan Science and Technology Agency (JST), Japan Agency for Medical Research and Development (AMED), EMBL, Wellcome, UK Research and Innovation (UKRI), US National Science Foundation (NSF), US National Institutes of Health (NIH), US Department of Energy (DOE)
<a href="#">ProteomeXchange Consortium</a>	UK, China, Japan, USA	The ProteomeXchange Consortium was established to provide globally coordinated standard data submission and dissemination pipelines involving the main proteomics repositories, and to encourage open data policies in the field.	EMBL, Wellcome, UK Research and Innovation (UKRI), US National Institutes of Health (NIH), US National Science Foundation (NSF), University of California, Japan Science and Technology Agency (JST), Chinese National Infrastructure for Protein Science, European Commission, Panorama Partners Program, Luxembourg National Research Fund
<a href="#">Rat Genome Database</a>	USA	The Rat Genome Database (RGD) was established in 1999 and rapidly became the premier site for genetic, genomic, phenotype, and disease-related data generated from rat research. In addition, RGD has expanded to include a large body of structured and standardized data for ten species (rat, mouse, human, chinchilla, bonobo, 13-lined ground squirrel, dog, pig, green monkey/vervet and naked mole-rat).	US National Institutes of Health (NIH)
<a href="#">Reactome</a>	Canada, UK, USA	REACTOME is an open-source, open access, manually curated and peer-reviewed pathway database. Our goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic and clinical research, genome analysis, modeling, systems biology and education.	US National Institutes of Health (NIH), EMBL, University of Toronto, European Commission
<a href="#">Rhea</a>	Switzerland	Rhea is an expert-curated knowledgebase of chemical and transport reactions of biological interest - and the standard for enzyme and transporter annotation in UniProtKB. Rhea uses the chemical dictionary ChEBI (Chemical Entities of Biological Interest) to describe reaction participants.	Swiss State Secretariat for Education, Research, and Innovation (SERI)

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2023
<a href="#">Saccharomyces Genome Database</a>	USA	The Saccharomyces Genome Database (SGD) provides comprehensive integrated biological information for the budding yeast <i>Saccharomyces cerevisiae</i> along with search and analysis tools to explore these data, enabling the discovery of functional relationships between sequence and gene products in fungi and higher organisms.	US National Institutes of Health (NIH)
<a href="#">SILVA</a>	Germany	SILVA provides comprehensive, quality checked and regularly updated datasets of aligned small (16S/18S, SSU) and large subunit (23S/28S, LSU) ribosomal RNA (rRNA) sequences for all three domains of life (Bacteria, Archaea and Eukarya).	Leibniz Institute DSMZ (DE), German Federal Ministry of Education and Research (BMBF), Max Planck Institute for Marine Microbiology, Gordon and Betty Moore Foundation, Deutsche Forschungsgemeinschaft (DE)
<a href="#">STRING</a>	Switzerland, Denmark, Germany	STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.	Swiss State Secretariat for Education, Research, and Innovation (SERI), University of Copenhagen, EMBL
<a href="#">UCSC Genome Browser</a>	USA, Germany, Japan	The UCSC Genome Browser is a web-based tool serving as a multi-powered microscope that allows researchers to view all 23 chromosomes of the human genome at any scale from a full chromosome down to an individual nucleotide. The browser integrates the work of countless scientists in laboratories worldwide, including work generated at UCSC, in an interactive, graphical display. The Browser also affords access to the genomes of more than one hundred other organisms.	US National Institutes of Health (NIH), California Institute for Regenerative Medicine, US Centre for Disease Control and Prevention (CDC)
<a href="#">UniProt</a>	UK, Switzerland, USA	UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information.	US National Institutes of Health (NIH), EMBL, UK Research and Innovation (UKRI), Swiss State Secretariat for Education, Research, and Innovation (SERI), US National Science Foundation (NSF)
<a href="#">VEuPathDB</a>	USA, UK	VEuPathDB provides access to diverse genomic and other large scale datasets related to eukaryotic pathogens and invertebrate vectors of infectious disease.	US National Institutes of Health (NIH), Bill and Melinda Gates Foundation, Wellcome
<a href="#">WormBase</a>	USA, Canada, UK	WormBase is an international consortium of biologists and computer scientists providing the research community with accurate, current, accessible information concerning the genetics, genomics and biology of <i>C. elegans</i> and related nematodes.	US National Institutes of Health (NIH), UK Research and Innovation (UKRI)



Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2023
<a href="#">ZFIN: The Zebrafish Information Network</a>	USA	The Zebrafish Information Network (ZFIN) is the database of genetic and genomic data for the zebrafish ( <i>Danio rerio</i> ) as a model organism. ZFIN provides a wide array of expertly curated, organized and cross-referenced zebrafish research data.	US National Institutes of Health (NIH)