

Global Core Biodata Resources

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2022
Alliance of Genome Resources	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)
BacDive	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)
BRENDA	Germany	European Commission, EMBL, UK Research and Innovation (UKRI)	German Federal Ministry of Education and Research (BMBF)
ChEBI	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)
ChEMBL	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 Trust, EMBL, UK Research and Innovation (UKRI)
CIViC: Clinical Interpretation of Variants in Cancer	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)
Clinical Genome Resource	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)
DNA Data Bank of Japan	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)
EcoCyc	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)

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2022
Ensembl	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 Trust, UK Research and Innovation (UKRI), European Commission, EMBL, US National Institutes of Health (NIH), Save the Tasmanian Devil Program Appeal
Europe PMC	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: https://europepmc.org/Funders/)
European Nucleotide Archive	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 Trust
FlyBase	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 Trust, UK Research and Innovation (UKRI)
GENCODE	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
Gene Ontology	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)
Global Biodiversity Information Facility	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 https://www.gbif.org/the-gbif-network), Danish Government, US National Science Foundation, European Commission, Bill & Melinda Gates Foundation
gnomAD	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 Trust, Michael J Fox Foundation

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2022
GWAS Catalog	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)
HGNC: HUGO Gene Nomenclature Committee	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 Trust, US National Institutes of Health (NIH)
InterPro	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 Trust, UK Research and Innovation (UKRI), EMBL, US National Institutes of Health (NIH)
MGD: Mouse Genome Database	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)
Orphadata Science	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
PANTHER	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)
PharmGKB: Pharmacogenomics Knowledgebase	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)
PomBase	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 Trust

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2022
Protein Data Bank	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 Trust, UK Research and Innovation (UKRI), US National Science Foundation (NSF), US National Institutes of Health (NIH), US Department of Energy (DOE)
ProteomeXchange Consortium	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 Trust, 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
Rat Genome Database	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)
Reactome	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
Rhea	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)
Saccharomyces Genome Database	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)

Name	Host Countries (Primary, Additional)	Overview	Funders, 2017-2022
STRING	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
UCSC Genome Browser	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)
UniProt	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)
VEuPathDB	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 Trust
WormBase	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)
ZFIN: The Zebrafish Information Network	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)