The European Bioinformatics Institute · Cambridge

Overview

EMBL-EBI

Welcome

Welcome to the European Bioinformatics Institute (EMBL-EBI), a global hub for big data in biology. We promote scientific progress by providing freely available data to the life-science research community, and by conducting exceptional research in computational biology.

At EMBL-EBI, we manage public life-science data on a very large scale, offering a rich resource of carefully curated information. We make our data, tools and infrastructure openly available to an increasingly data-driven scientific community, adjusting to the changing needs of our users, researchers, trainees and industry partners. This proactive approach allows us to deliver relevant, up-to-date data and tools to the millions of scientists who depend on our services.

We are a founding member of ELIXIR, the European infrastructure for biological information, and are central to global efforts to exchange information, set standards, develop new methods and curate complex information. Our core databases are produced in collaboration with other world leaders including the National Center for Biotechnology Information in the US, the National Institute of Genetics in Japan, SIB Swiss Institute of Bioinformatics and the Wellcome Sanger Institute in the UK.

We are also a world leader in computational biology research, and are well integrated with experimental and computational groups on all EMBL sites. Our research programme is highly collaborative and interdisciplinary, regularly producing high-impact works on sequence and structural alignment, genome analysis, basic biological breakthroughs, algorithms and methods of widespread importance.

EMBL-EBI is an international treaty organisation, and we serve the global scientific community across academia and industry. As new technologies continue to accelerate radical changes in medicine and agriculture, our member states know they can rely on us to share our expertise in technology, training and data analysis.

We warmly encourage you explore what EMBL-EBI has to offer, and to get in touch if you have any questions.

Sincerely,

Dr Rolf Apweiler, Director

Dr Ewan Birney, Director

About us

- ◎ EMBL-EBI is a global leader in bioinformatics: the science of storing, analysing and sharing large biological datasets.

- ◎ By sharing our expertise and through collaboration, we help researchers realise the potential of 'big data', enhancing their ability to exploit complex information to make discoveries that benefit mankind.

- ◎ We are a non-profit, intergovernmental organisation funded by EMBL member states.

- ◎ Our 650 staff represent 66 nationalities and we welcome a regular stream of visiting scientists throughout the year.

- ◎ We are located on the Wellcome Genome Campus in Hinxton, Cambridge, in the United Kingdom.

Our mission

- To provide freely available data and bioinformatics services to all facets of the scientific community

- To contribute to the advancement of biology through basic investigator-driven research

- To provide advanced bioinformatics training to scientists at all levels

- To help disseminate cutting-edge technologies to industry

- To support, as an ELIXIR Node, the coordination of biological data provision throughout Europe

The European Molecular Biology Laboratory (EMBL)

We are part of EMBL, Europe's flagship laboratory for the life sciences. Research at EMBL is conducted by over 80 independent groups covering the full spectrum of molecular biology.

EMBL's main laboratory is located in Heidelberg, Germany. EMBL's other sites specialise in different disciplines: bioinformatics at EMBL-EBI in Hinxton, Cambridge, UK; neurobiology and epigenetics in Rome, Italy; structural biology in Grenoble, France, and Hamburg, Germany; and tissue biology in Barcelona, Spain.

Leadership

The Directors of EMBL-EBI are Dr Rolf Apweiler and Professor Ewan Birney.

EMBL's Director General is Professor Iain Mattaj. Its governing body, EMBL Council, comprises representatives of all member and associate member states.

The EMBL Council appoints a Scientific Advisory Committee (SAC), which plays an important role in guiding the programme of the laboratory.

EMBL-EBI bioinformatics services largely represent global collaborations, and these are supported by individual scientific advisory committees.

How EMBL is funded

EMBL is an intergovernmental organisation funded by public research monies from over 20 member states and associate member states.

EMBL member states: Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Luxembourg, Malta, Montenegro, the Netherlands, Norway, Portugal, Slovakia, Spain, Sweden, Switzerland and the United Kingdom.

EMBL associate member states: Argentina and Australia.

EMBL prospect member states: Poland and Lithuania.

How EMBL-EBI is funded

The largest part of EMBL-EBI funding comes from the governments of EMBL's member states.

The global importance of our work is reflected in the fact that we also attract significant funds from the European Commission, the US National Institutes of Health, the Wellcome Trust, the Research Councils and our industry partners.

The Biotechnology and Biological Sciences Research Council (BBSRC), under the umbrella of UK Research and Innovation (UKRI), has provided substantial capital funding to further develop our robust compute infrastructure and to expand our activities in our South Building.

Data deposition

Nucleotides
Transcripts
Proteins
Complexes
Domains
Structures
Small molecules
Pathways
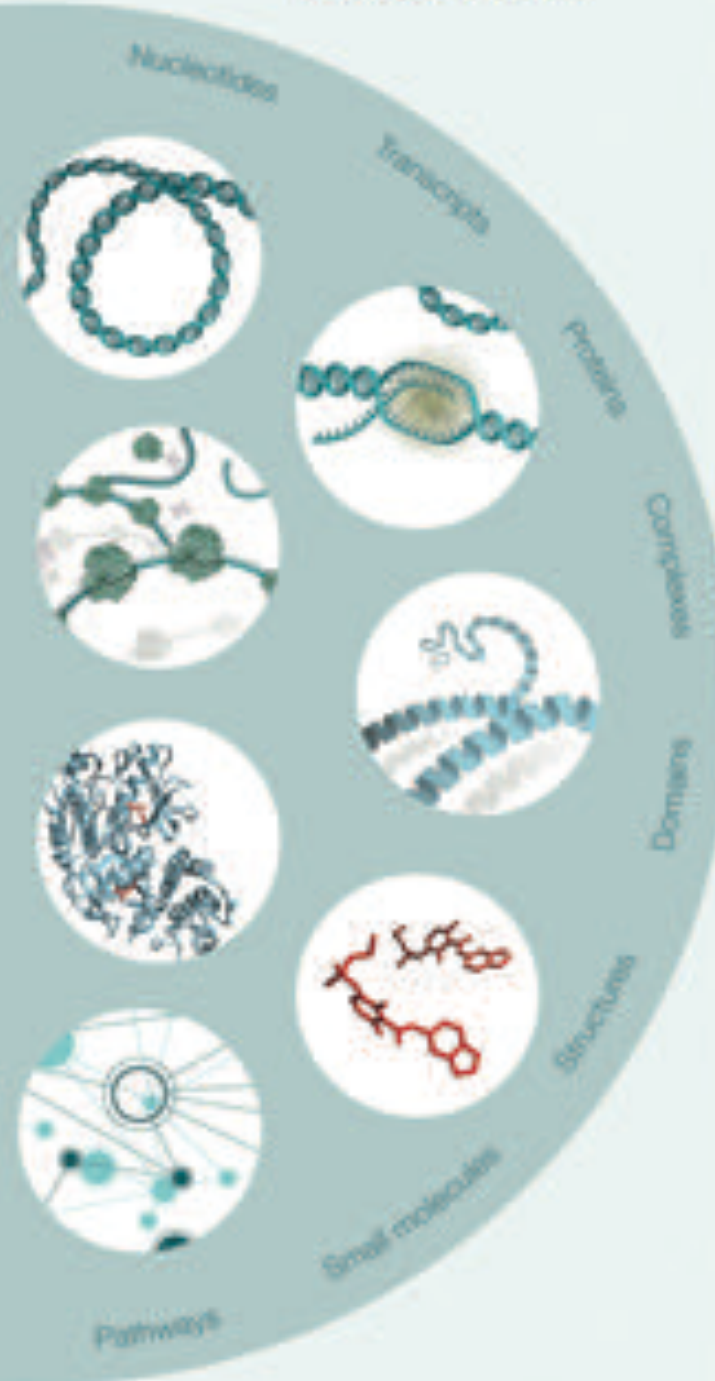
Integration

Human beings
and other organisms

Tissues & organs

Cells

Biobanks

Translation

Populations

New treatments

Disease prevention

Early diagnosis

From molecules to medicine

Biology has changed radically as disruptive technologies such as next-generation genome sequencing, single-cell sequencing and high-performance computing have matured. As new data types come online, new challenges and opportunities emerge that demand a critical mass of bioinformatics expertise.

EMBL-EBI is situated on the Wellcome Genome Campus in the UK, one of the largest concentrations of bioinformatics and genomics expertise in the world. We serve as anchor partners in large-scale, cutting-edge research endeavours, often leading in data management, sharing, security and analysis.

Our expertise is crucial to the scaling up of genome sequencing projects that focus on topics such as cancer, rare diseases, marine systems and staple crops, and to establishing common vocabularies that bring research communities together.

We are a trusted advisor to governments seeking to incorporate genomics into healthcare, and have a growing set of clinical collaborations, including a strategic relationship with the School of Clinical Medicine at the University of Cambridge.

Selected collaborations

COMPARE Project

Speeding up the detection, analysis and comparison of crucial epidemiological information using genome technology.

International Cancer
Genome Consortium

Obtaining a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types of clinical and societal importance.

Global Alliance for Genomics and Health
Developing and testing the data standards needed to bring genomics into the clinic and enable precision medicine.

Human Cell Atlas
Charting the specific genetic properties of all human cells, across all tissues and organs, and building a new reference map of the healthy human body.

HipSci

Creating a UK national stem cell resource and using it to discover how genomic variation impacts cellular phenotype.

International Mouse Phenotyping
Consortium

The first truly comprehensive, functional catalogue of a mammalian genome.

Designing Future Wheat
A BBSRC-funded programme that aims to develop new wheat germplasm containing the next generation of key traits, with the aim of supporting sustainable and productive agriculture.

Data resources

At EMBL-EBI, we see big data as a critical tool that can accelerate research and development.

We maintain the world's most comprehensive range of freely available biological databases, and are constantly working to provide opportunities for scientists to make the best possible use of public and proprietary data.

We help public- and private-sector researchers focus on innovative work by providing access to our sophisticated and large-scale data infrastructure.

We provide numerous tools and methods that enable researchers to perform large-scale analysis, either using our services or by downloading resources for local use.

A powerful driver for life science

Big data has become a driver in pharmaceutical, biotech and agricultural R&D, largely thanks to low-cost DNA sequencing and other high-throughput technologies. Drawing on large volumes of public and proprietary information allows commercial and academic researchers to gain new insights into challenges in health and the environment.

As a public, internationally funded organisation, we provide a sustainable platform for managing problems with global dimensions. For example, public health experts looking into a disease outbreak can use DNA sequencing to identify small variations between strains of the virus, track transmission and identify the source. This same technology can be used to sequence hospital environments, providing a sensitive tool allowing staff to act quickly to contain a problem at the earliest possible stage. EMBL-EBI provides infrastructure to house this information and make it available on demand.

We make molecular data and infrastructure freely available so that researchers in all sectors can maximise the potential of life-science research. The services we deliver enrich research and development by:

- ◎ Optimising efficiency, saving time and improving productivity

- ◎ Enabling the re-use of data in many different disciplines

- ◎ Providing results in context for easier interpretation

- ◎ Maximising the potential of basic research

- ◎ Empowering researchers and promoting innovation

Principles of data service provision

Accessibility

Our data and tools are freely available, without restriction, the only exception being potentially identifiable human genetic information, for which access depends on research consent agreements.

Quality

Our databases are enhanced through annotation: information is added to the original data to provide context and help with interpretation. Much of our annotation is performed by highly qualified biologists, with automated annotation subjected to rigorous quality control.

Portability

Many of our datasets, and in some cases whole software systems, are made available for download from the EMBL-EBI website. We share data to ensure that our resources are comprehensive and up to date, and work with publishers to ensure biological data are shared in public repositories and cross-referenced in the relevant publication.

Compatibility

We are a world leader in promoting the adoption of standards in bioinformatics.

Data resources

Genes, genomes & variation

1000 Genomes
A deep catalogue of shared human genetic variation in population groups worldwide.

Database of Genomic Variants archive
A repository for the archiving, accessioning and distribution of publicly available genomic structural variants, in all species.

MGnify
A resource for the analysis and archiving of metagenomic data.

European Genome–phenome Archive
A service for the permanent archiving and sharing of all types of personally identifiable genetic and phenotypic data resulting from biomedical research projects.

European Variation Archive
An open-access database of all types of genetic variation data, from all species.

European Nucleotide Archive
A comprehensive archive of submitted nucleotide sequence-read, assembly and functional annotation data.

Ensembl
High-quality, integrated annotation on vertebrate genomes within a consistent and accessible infrastructure.

Ensembl Genomes
An integrating portal for genome-scale data from non-veterbrate species, including bacteria, protists, fungi, plants and invertebrate metazoa.

RNAcentral
A database of non-coding RNA sequences.

Expression

ArrayExpress
A database of functional genomics experiments including microarray and RNAseq expression data typically related to publications.

Expression Atlas
An added-value database that shows which genes are expressed under which conditions, and how expression differs between conditions.

PRIDE
An archive of protein-expression data determined by mass spectrometry.

MetaboLights
A cross-species, cross-application, open-access, open-submission archive and reference database for metabolomics.

Proteins: sequences, families, domains & motifs

InterPro
A database for the classification of proteins into families, domains and conserved sites.

Pfam
A database of hidden Markov models and alignments to describe conserved protein families and domains.

UniProt
The Universal Protein Resource – a comprehensive resource for protein sequence and functional annotation data.

Molecular & cellular structures

Electron Microscopy Data Bank
Electron microscopy density maps of macromolecular complexes and subcellular structures. EMDB covers a variety of techniques, including single-particle analysis, electron tomography, and electron (2D) crystallography.

Protein Data Bank in Europe

The European resource for the collection, organisation and dissemination of 3D structural data (from the Protein Data Bank and the Electron Microscopy Data Bank) on biological macromolecules and their complexes.

Chemical biology

ChEBI

Chemical Entities of Biological Interest – Reference chemical structures, nomenclature and ontological classification.

ChEMBL

An open-data resource of binding, functional and ADMET bioactivity data. SureChEMBL provides free access to chemical data extracted from the patent literature.

SureChEMBL

Chemical structure data from the full text literature and images of patents, helping researchers check whether a newly developed drug or other product is actually novel.

Pathways & systems

BioModels

Search and retrieve published mathematical models of biological interest. Models are annotated and linked to relevant data resources.

BioSamples

Explore information about reference samples (e.g. Coriell Cell lines) and sample data from ArrayExpress, the ENA and PRIDE, with links to assays.

Enzyme Portal

Integrated enzyme data from EMBL-EBI resources. Integrated functional, sequence, nomenclature, substrate, product and cofactor data for enzymes.

IntAct

An open-source database system and analysis tools for molecular interaction data. All interactions are derived from literature curation or direct user submissions.

Reactome

Navigate a map of human biological pathways, ranging from metabolic processes to hormonal signalling.

Literature

BioStudies

Descriptions of biological studies, links to data from these studies, and data that do not fit in the structured archives at EMBL-EBI. For example, BioStudies can hold supplementary material linked to published papers.

Europe PubMed Central

Free access to life-science literature resources including: PubMed abstracts, full-text articles, patent abstracts, NHS clinical guidelines, Agricola records and more. Supplemented with Chinese Biological Abstracts and the Citeseer database.

Research

EMBL-EBI has been leading computational biology research since its inception in 1994, with deep involvement in large-scale endeavours such as the Human and Mouse Genome Projects and ENCODE, and publishing seminal works in structural biology, chemical biology and single-cell genomics.

As part of EMBL, we are well-integrated with experimental and computational research programmes throughout Europe. Most of our researchers collaborate closely with experimentalists and even generate experimental data themselves.

Selected research

Ageing: cell coordination breakdown

Study showed that immune cells in older tissues lack coordination and exhibit much more variability in gene expression compared with their younger counterparts. Martinez-Jimenez CP, et al (2017). Science

Social Genetic Effects

The health of individual mice is influenced by the genetic makeup of their partners, indicating that research into genetics and disease should include the genotypes of both individuals and their partners. Baud et al. (2017). PLOS Genetics

Genomics reboots deep learning

A new deep-learning method leverages 'deep neural networks', multi-layered machine-learning models inspired by the brain, to gain new insights into health and disease. Angermueller C, et al. (2017). Genome biology

Using DNA to store digital information

EMBL-EBI researchers created a way to store data in the form of DNA – a material that lasts for tens of thousands of years with minimal intervention. The new, error-free method makes it possible to store at least 100 million hours of high-definition video in about a cup of DNA. Goldman N, et al. (2013). Nature

EMBL-EBI Research Groups

Bateman group
Analysis of protein and
RNA sequence
www.ebi.ac.uk/research/bateman

Beltrao group
Evolution of cellular networks
www.ebi.ac.uk/research/beltrao

Birney group
Sequence algorithms and
intra-species variation
www.ebi.ac.uk/research/birney

Brazma group
Functional genomics research
www.ebi.ac.uk/research/brazma

Finn group
Sequence families
www.ebi.ac.uk/research/finn

Flicek group
Evolution of transcriptional
regulation
www.ebi.ac.uk/research/flicek

Gerstung group
Computational cancer biology
www.ebi.ac.uk/research/gerstung

Goldman group
Evolutionary tools for
genomic analysis
www.ebi.ac.uk/research/goldman

Iqbal group
Small molecule metabolism in
biological systems
www.ebi.ac.uk/research/iqbal

Leach group
Chemistry services
www.ebi.ac.uk/research/leach

Marioni group
Computational and
evolutionary genomics
www.ebi.ac.uk/research/marioni

Petsalaki group
Small molecule metabolism in
biological systems
www.ebi.ac.uk/research/petsalaki

Stegle group
Statistical genomics and
systems genetics
www.ebi.ac.uk/research/stegle

Thornton group
Proteins: structure,
function and evolution
www.ebi.ac.uk/research/thornton

Uhlman group
Bioimage analysis
www.ebi.ac.uk/research/uhlman

Zerbino group
Genome analysis
www.ebi.ac.uk/research/zerbino

Training

Data analysis is a major bottleneck in life-science research, and Europe's growing, diversifying research community needs ongoing training to adapt in a rapidly changing technical environment.

EMBL-EBI helps researchers get the most out of publicly available data and tools in an extensive user-training programme.

Our courses and free online tutorials help scientists familiarise themselves with the data and services most relevant to their work.

This helps them develop the skills and confidence they need to make the most of the vast amounts of biological data in the public domain.

The Bioinformatics Training Programme

The EMBL-EBI Training Programme, part of the EMBL International Centre for Advanced Training, helps users get to grips with their data online, in courses on site and in events held at host organisations throughout the world.

We emphasise demonstration, discussion and practical exercises. Our hands-on courses, conducted in state-of-the-art facilities, are led by experts from EMBL-EBI and from other leading research organisations throughout the world.

Our online learning programme, Train online, is based on the materials created for these courses.
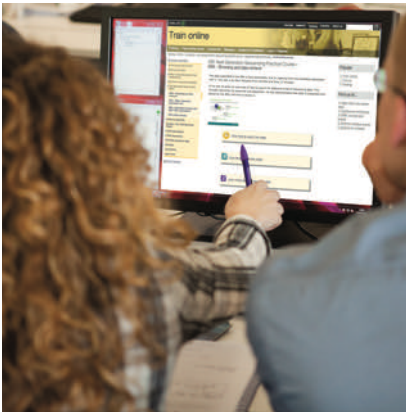
Online training

Our free Train online resource offers interactive courses and is ideally suited for on-the-job training.

www.ebi.ac.uk/training/online

We also offer a freely accessible collection of webinars for users of bioinformatics tools and services, at all levels.

www.ebi.ac.uk/training/webinar

Training hosted offsite

Experts from EMBL-EBI are often invited to run courses at host institutes throughout the world, which have specific R&D needs. In addition, many of our service teams run training events offsite that are dedicated to a single resource.

www.ebi.ac.uk/training/roadshow

Training offered on site

The EMBL-EBI Training Programme offers hands-on training in custom-built, state-of-the-art facilities. Topics range from courses for experimental biologists to more specialised workshops for computational biologists working in both the public and private sectors.

www.ebi.ac.uk/training

Supporting industry

The EMBL-EBI Industry Programme

The EMBL-EBI Industry Programme actively promotes pre-competitive collaboration, open-source software and informatics standards to improve efficiency and reduce costs.

The programme is a forum for interaction and knowledge exchange for those working at the forefront of large-scale, commercial bioinformatics.

Since 1996, our Industry Programme has been helping companies make the most of advances in bioinformatics and cheminformatics. Our industry collaborations include metagenomics, single-cell genomics and using cheminformatics to understand diabetes.

Our members lend their insights and expertise to help define and shape the services we provide, and meet regularly to discuss shared challenges and identify topics for training workshops.

The strategic focus is the development of resources and services that will benefit our members and our wider stakeholder communities.

Industry Programme members

Pharmaceutical and diagnostic

Agri-food and personal care

Industry collaboration

An important part of EMBL-EBI's mission is to disseminate cutting-edge technologies to industry. A healthy percentage of our users is engaged in industrial R&D, and our services are constantly evolving to reflect the rapidly changing needs of this crucial sector. We do this with the help of our Technology Transfer office, EMBLEM.

We welcome new collaborations with industry at all levels, from large pharma and agri-food companies, to biotech and small and medium-sized enterprises (SMEs).

Open Targets for new medicines

In drug discovery, over half of new medicines fail in late-stage development, at enormous cost to companies. One reason for this is insufficient knowledge about the nature of a gene or protein that the medicine is meant to target. Companies decide to pursue a biological target based on analyses of diverse data and software applications – but the lack of systematic, reliable approaches has made target validation the Achilles heel of the industry.

GSK, Biogen, Takeda, Celgene, EMBL-EBI and the Wellcome Sanger Institute are taking a fresh approach to the problem. Open Targets, based in our South Building, tackles the underlying biology systematically, using computational biology to verify – early in the drug discovery process – whether acting on a target is likely to provide therapeutic benefit.

Open Targets aims to make it easier for researchers to link a gene target to a disease by collating, integrating and scoring the available evidence.

 It is committed to sharing its results openly, providing clarity and informing important decisions that impact drug-discovery investment.

www.opentargets.org

European coordination

Europe has always been at the forefront of bioinformatics research, and as we move towards a single European Research Area there is a greater need than ever to consolidate our efforts.

EMBL-EBI hosts the Technical Hub of ELIXIR, the European infrastructure for biological information. The purpose of ELIXIR is to support research and its translation to medicine and the environment, the bio-industries and society.

EMBL-EBI participates in several ELIXIR projects, including CORBEL, which seeks to harmonise user access to life-science resources, and EXCELERATE, an initiative to fast-track the implementation of life-science data infrastructure.

www.elixir-europe.org

Working on the Genome Campus

The Wellcome Genome Campus is home to some of the world's most advanced institutes working at the interface of genomics and computational biology.

Working here gives you an opportunity to apply your skills and energy for the greater good, in a friendly and collegial atmosphere. Our diverse, exceptional scientific community creates a culture and environment that fosters creativity.

EMBL-EBI is curiosity-driven, community-oriented and international. Our teams are flexible and collaborative, and the projects we take on require all kinds of people and working styles. We are proud to be an equal-opportunity employer.

We hold open days and welcome anyone who is interested in what we do, or who would like to pursue a career in bioinformatics.

www.ebi.ac.uk/about/jobs

How to find us

EMBL-EBI is situated on the Wellcome Genome Campus alongside the village of Hinxton, 10 miles (16 km) south of Cambridge and 30 km from London Stansted airport. The campus sits between two local railway stations (Whittlesford Parkway and Great Chesterford) and close to Junction 10 of the M11 motorway.

You can find detailed directions and local accommodation information on our website.