

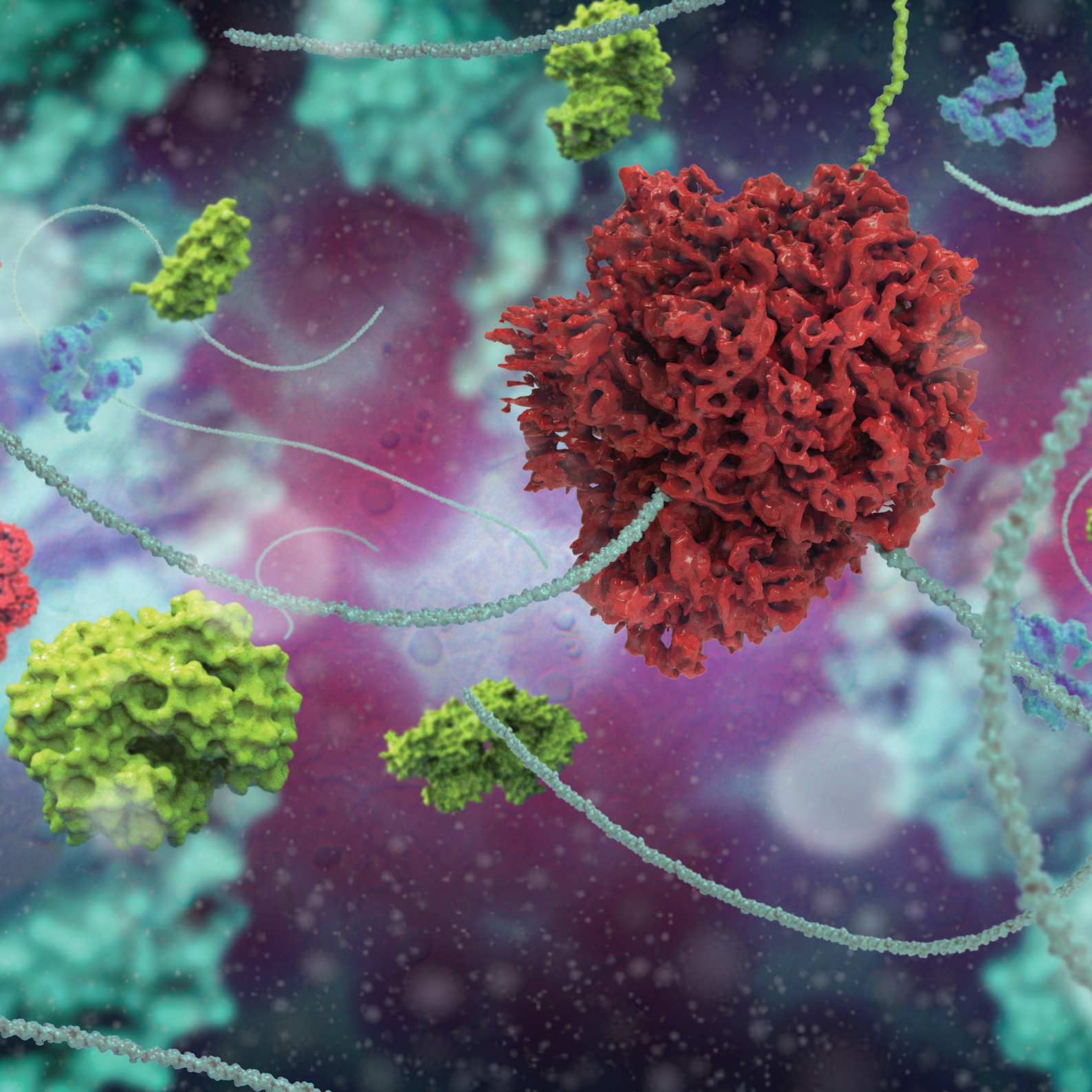


The European Bioinformatics Institute

Innovation & Translation

EMBL-EBI





A Powerful Driver for Research

Big data has become a driver in pharmaceutical, biotech and agricultural R&D, thanks to low-cost DNA sequencing and other high-throughput technologies. By drawing on the large volumes of public and proprietary information produced in commercial and academic research, scientists are rapidly gaining valuable insights into major challenges in health, agriculture and the environment.

The European Bioinformatics Institute (EMBL-EBI) provides services that keep Europe at the forefront of research and development. We make data and infrastructure freely available to everyone in order to spur innovation and maximise the potential of biomedical and life-science research.

We serve as an anchor partner in large-scale, cutting-edge research endeavours, often leading on data management, sharing, security and analysis. Our expertise is crucial to scaling up genome sequencing projects that focus on cancer, rare diseases, marine systems and staple crops, and to establishing common vocabularies that bring research communities together.

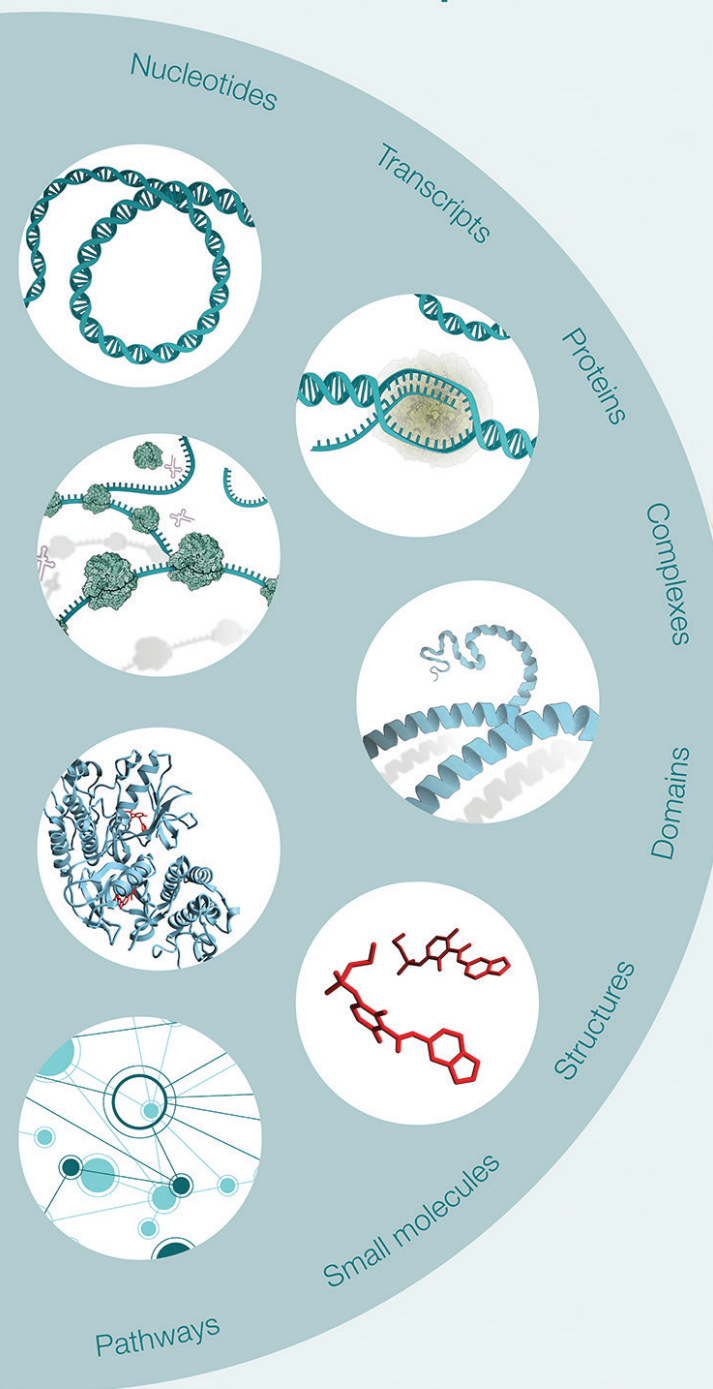
Our data services are designed to enrich research and development by:

- Enabling the re-use of data in many different disciplines;
- Providing the context for accurate interpretation of complex data;
- Optimising efficiency, saving time and improving productivity;
- Empowering researchers and promoting innovation;
- Maximising the potential of basic research.

We are part of EMBL, which spans six sites in Germany, France, Italy, Spain and the UK. We encourage the transfer of knowledge and innovative technology to industry through pre-competitive collaboration, licensing, service provision and creation of spin-out companies.

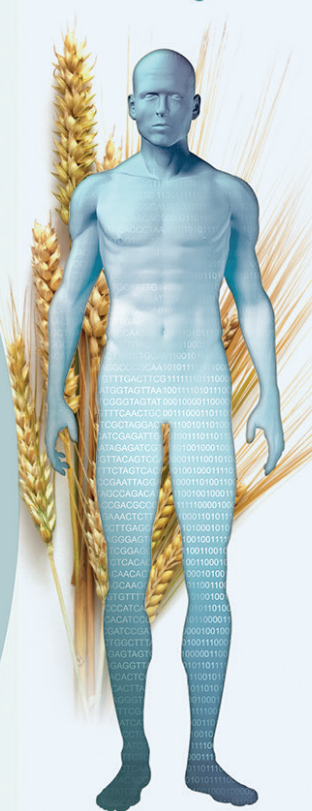
Located on the Wellcome Genome Campus just south of Cambridge, we are at the centre of one of the highest concentrations of technical and scientific expertise in the world. With around 600 members of staff and expertise in all aspects of computational biology, we are at the heart of bioinformatics in Europe.

Data deposition



Integration

Human beings
and other organisms



Tissues & organs



Cells



Biobanks

Translation







Populations









New treatments



Disease prevention

GLUCOSE: read at exactly 30 seconds						
Negative	%	1/10	1/4	1/2	1	2 or more
						
	mg/dl	100	250	500	1000	2000+

KETONE: read at exactly 15 seconds						
Negative	Trace	Small	Moderate	Large		
						
	mg/dl	5	15	40	80	160

Early diagnosis



From Molecules to Medicine

A genome is the complete set of DNA in any organism. It holds the key to greater understanding of our development, how and why we differ, and what makes us susceptible to diseases. We can begin to understand what is happening in individual patients by comparing their DNA with 'reference' genomes, and combining that information with other types of data such as the level of proteins found in blood.

EMBL-EBI provides free data and tools that help clinicians investigate whether their patient has an increased risk for disease. We are always working to expand these resources so that scientists everywhere can benefit from technological advances such as next-generation sequencing, single-cell sequencing and high-performance computing.

We are a pivotal partner in international projects that seek to transform biomedicine, sharing our data-management and bioinformatics expertise. We make the results of these global endeavours reusable by future generations by integrating them into our public data services. This paves the way for a shift in healthcare towards genomics and personalised medicine.

EMBL-EBI: A Key Partner in Biomedical and Public Health Projects

- **Cancer:** The International Cancer Genome Consortium is obtaining a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types of clinical and societal importance.
- **Stem cells:** HipSci is a UK national stem cell initiative for discovering how genomic variation impacts cellular phenotype.
- **Epigenetics:** BLUEPRINT offers reference epigenomes of distinct types of blood cells from healthy individuals and of their malignant leukaemic counterparts.
- **Toxicology:** eTOX is developing innovative strategies and novel software tools to predict the toxicological profiles of small molecules in the early stages of drug development.
- **Health Records:** EHR4CR is designing a scalable, cost-effective approach to interoperability between electronic health record systems and clinical research.
- **Mapping Genotype to Phenotype:** The International Mouse Phenotyping Consortium provides free, unrestricted access to primary and secondary data from a comprehensive, functional catalogue of a mammalian genome.
- **Infectious Disease:** COMPARE is speeding up the detection of and response to infectious disease outbreaks using genome technology and rapid, secure data sharing. The ZIKA Alliance is investigating the clinical, fundamental, environmental and social aspects of ZIKA infection.

Scientific Data Services

At EMBL-EBI, we see data as a critical tool that can accelerate research. We maintain the world's most comprehensive range of biological databases, and help scientists everywhere turn information into knowledge.

Genes, genomes & variation

European Nucleotide Archive

The comprehensive archive of submitted nucleotide sequence-read, assembly and functional annotation data.

EBI Metagenomics

For the analysis and archiving of metagenomics data from environmental samples.

European Genome-phenome Archive

For the archiving and sharing of all types of personally identifiable genetic and phenotypic data resulting from biomedical research projects.

European Variation Archive

For access to all types of genetic variation data, from all species.

1000 Genomes

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

Ensembl

High-quality, integrated annotation on the genomes of vertebrate species within an accessible infrastructure offering advanced tools for analysis.

Ensembl Genomes

For access to genome-scale data from plant, fungal, bacterial, protist and metazoan species.

GWAS Catalog

A quality-controlled, manually curated, literature-derived collection of all published genome-wide association studies.

RNACentral

For access to non-coding RNA sequence data from an international consortium of RNA resources.

Expression

ArrayExpress

For access to data from functional genomics experiments, including microarray and RNAseq expression data.

Expression Atlas

For exploring which genes or proteins are expressed under different conditions, and comparing diseased and healthy states.

MetaboLights

For metabolomics experiments and derived information.

PRIDE

For access to protein-expression data determined by mass spectrometry.

Protein sequences

UniProt

A comprehensive, foundational resource for protein sequence and functional annotation data.

InterPro

Classify proteins into families and predict the presence of domains and functionally important sites.

Pfam

For access to hidden Markov models and alignments to describe conserved protein families and domains.

Molecular & cellular structures

Protein Data Bank in Europe

For the collection, organisation and dissemination of 3D structural data on biological macromolecules and their complexes.

Electron Microscopy Data Bank

Access and analyse electron microscopy density maps of complexes and subcellular structures.

Chemical biology

ChEBI

Reference chemical structures, nomenclature and ontological classification.

ChEMBL

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

Pathways & systems

IntAct

For sharing and analysing molecular interaction data derived from literature curation and user submissions.

Reactome

An interactive map of human biological pathways, from metabolic processes to hormonal signalling.

BioSamples

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

Enzyme Portal

For access to all functional, sequence, nomenclature, substrate, product and cofactor data for enzymes in EMBL-EBI resources.

Literature

Europe PubMed Central

Access and mine full-text, open life-science literature linked to molecular data resources.

BioStudies

For studies and data that do not fit in the structured archives, e. g. supplementary material linked to published papers, or unconventional datasets not linked to any publication.

Train online

Free, web-based bioinformatics tutorials and webinars.

Research

Our research produces high-impact, novel methods for analysing genomes on a large scale, identifying carcinogens, combining diverse, emerging data types and separating the signal from the noise. Our research themes in 2017 span genomic analysis to integrative systems biology, including:

- Sophisticated, multi-dimensional statistical models linking genotypes and phenotypes
- Evolutionary regulatory genomics
- Single-cell genomics
- Cancer genomics, including clinical impact
- Large-scale imaging genetics
- Infectious diseases
- Evolution of post-translational modifications across species
- Structural biology

Complex, large-scale genome analysis made easier

A new statistical approach to studying the effect of genetic variations on different traits makes it possible to perform genetic correlations of up to 500,000 individuals – and many traits – at the same time. This innovation has applications across many domains. Casale FP, et al. (2015) *Nature Methods*

Genetic makeup of ‘room-mate’ influences health

Healing and anxiety are influenced by the genetics of one’s social partners, according to new research from EMBL-EBI. Baud et al. (2017) *PLOS Genetics*

Breast cancer breakthrough

The largest-ever study to sequence the whole genomes of breast cancers uncovered five new genes associated with the disease and 13 new mutational signatures that influence tumour development. The findings provide insights into the causes of breast tumours and demonstrate that breast-cancer genomes are highly individual. Morganella et al. (2016) *Nature Communications*

Cancer cell lines predict drug response

Patient-derived cancer cell lines harbour most of the same genetic changes found in patients' tumours, and could be used to learn how tumours are likely to respond to new drugs. The findings will help increase the success rate for developing more personalised cancer treatments. Iorio et al. (2016) *Cell*

Allergy: the price of immunity?

Molecular similarities between proteins in multicellular parasites and proteins that cause allergy were identified for the first time by EMBL-EBI researchers. The findings help demonstrate the evolutionary basis for allergy. Tyagi et al. (2015) *PLoS Computational Biology*

Towards an expression atlas for an entire brain

A new method for expression analysis, developed at EMBL Heidelberg and EMBL-EBI in Hinxton, increases the number of markers indicating specific cell types within a tissue by matching quantitative and spatial data. This makes it possible to identify wider gene-expression patterns within cell types, in different individuals and even across species. Achim K, et al. (2015) *Nature Biotechnology*

Exploring the universe of biochemical reactions

EC-BLAST, a new tool created by EMBL-EBI researchers, allows users to quickly compare the functions of thousands of catalysts at once. The software makes it easier for biotechnology researchers to develop novel enzymes. Rahman SA, et al. (2014) *Nature Methods*

Periodic table of protein complexes

A new Periodic Table of Protein Complexes provides a valuable tool for research into evolution and protein engineering. The Table enables a unified way to classify and visualise protein complexes, and anticipate new ones. Ahnert SE, et al. (2015) *Science*

About Our Impact

We asked our users
in an annual survey*:

“To what extent do EMBL-EBI services...”

reduce the time required
to find relevant data

90%

enable you to go ahead
with research you could
not otherwise have
pursued

79%

improve the quality of
data used within your
research

75%

reduce the financial cost
of exploiting relevant
data

65%

*1293 survey respondents in 2016

Essential to Research

72%

In our 2015 survey,
72% of respondents
agreed that EMBL-
EBI data and
resources were
“essential to their
research”

Training

350 000

In 2016 our online
training platform was
accessed by over
350 000 unique IP
addresses. Our staff
contributed to 350+
events, including
courses on site, off-
site workshops and
webinars.

Daily Usage

27 million

There are around 27
million requests to the
EMBL-EBI website
every day. A 2015
impact report showed
that our users spend
**20% of research
time** working with
EMBL-EBI data.

Collaboration

62 countries

In 2016, EMBL-EBI
had 186 grants funded
jointly with researchers
and institutes in 62
countries.

*(Does not include grants
for which EMBL-EBI was
the sole recipient)*

Rising Demand

120
Petabytes

To accomodate the
rapid growth in data and
demand, EMBL-EBI has
a total of 120 petabytes of
raw storage capacity.

Every month, scientists
at 3.2 million unique IP
addresses use
our services.

Publications

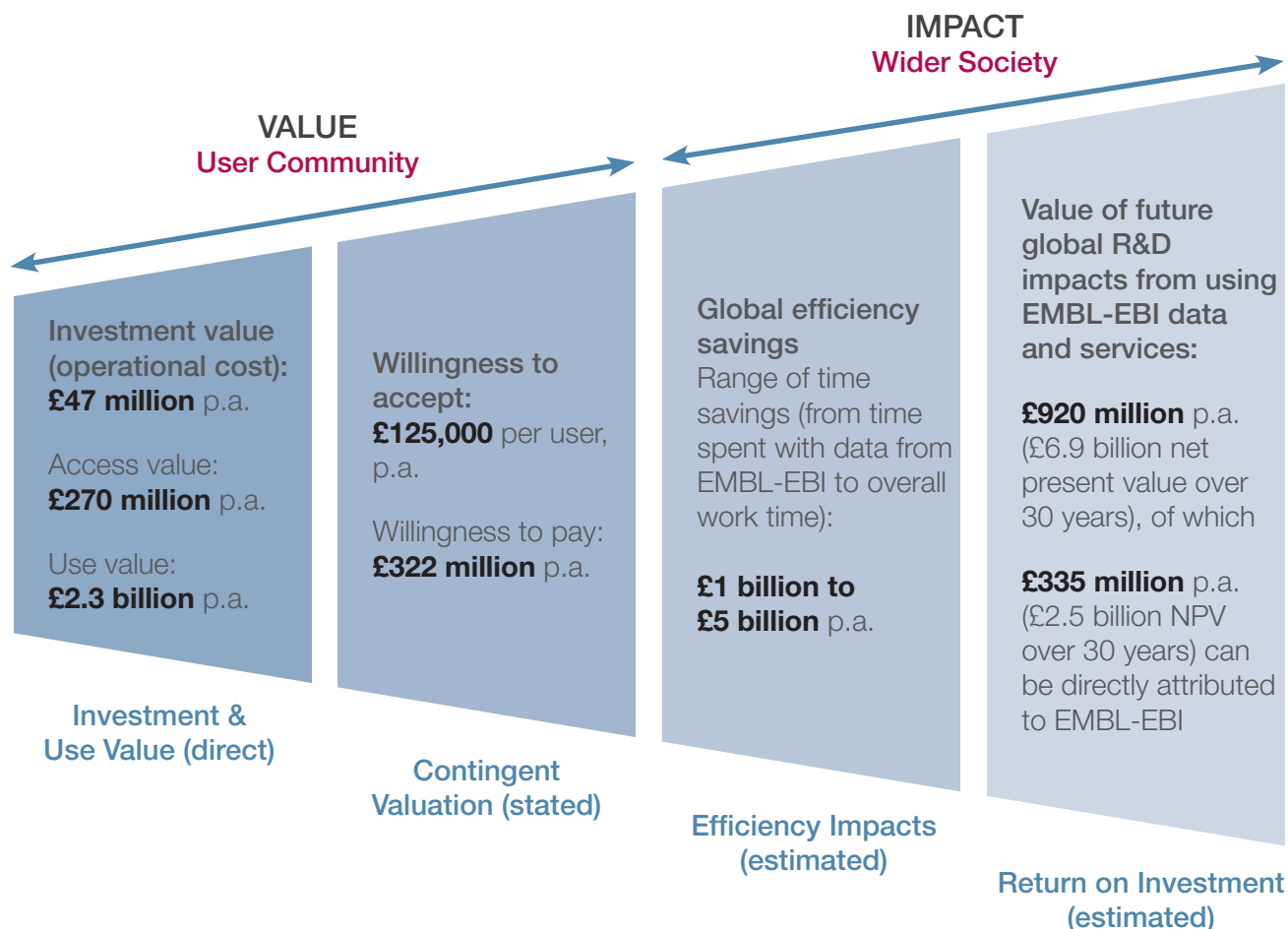
255 papers

EMBL-EBI scientists
published 255
papers in 2016 with
collaborators in
48 countries. Co-
publications with
industry historically
account for 15% of
papers published
annually.

The Economics

In 2015 an independent consultancy, Charles Beagrie Ltd., carried out an analysis of EMBL-EBI, employing a range of economic approaches to build a picture of direct user community effects to the wider commercial, healthcare and research communities. Their findings, summarised below, show that EMBL-EBI services are utilised widely and valued highly by the scientific community.

The unique nature of the institute's services is reflected in the results of a large-scale user survey, in which **45% of respondents** indicated they “could not have found the data they needed anywhere else, nor could they have recreated it”.



Turning Information Into Knowledge

A Critical Tool for Discovery

ChEMBL, an EMBL-EBI data resource, pulls together a vast amount of information about the effects of chemical compounds on living systems. It links data about chemical compounds, small molecules and biological targets, and is used routinely by researchers in the public and private sector.

ChEMBL enables the discovery of new treatments that benefit human, animal and environmental health, and promotes agricultural productivity.

“Without crop protection compounds, 40% of the world’s food would not exist. Our scientists use ChEMBL to support projects in our research towards innovative new products... People at EMBL-EBI do a fantastic job in making a vast amount of data of different types openly available to researchers, and without EMBL-EBI resources in general, I’m sure life science research would be greatly hindered.” - Mark Forster, Syngenta

The Gold Standard for Protein Data

Proteins—including hormones, antibodies and enzymes—are the functional units required to sustain life. They are composed of 20 amino acids, which can be arranged to create millions of different proteins, each one with a specific function.

EMBL-EBI leads the development of **UniProt**, the global hub for data and information about proteins. UniProt is a vital tool for research, as it unifies the available data for a given protein and adding expert annotation. UniProt is used by hundreds of thousands of researchers every month to turn information into knowledge, and knowledge into discovery.

UniProtKB has been cited over 15 000 times in Class A & C IPCR patents published since 2000, and over 10 000 times in scientific papers covering a diverse range of topics, including human disease.



Supporting Research into Cancer and Rare Diseases

The genomes of many organisms—including humans—have been fully sequenced, and their reference datasets are held in the Ensembl genome explorer at EMBL-EBI.

Ensembl offers clinically relevant tools including the **Variant Effect Predictor** (VEP), a powerful software package for analysing human variation and predicting the potential effects of genetic variants.

The VEP is used in research and clinical diagnosis of cancer and rare diseases, where strong links have been established between changes in the genome and development of a disease.

“Illumina’s VariantStudio talks to an annotation tool which has VEP at its core... Because of VEP’s superior quality and accuracy, its users are able to catch some edge cases where annotation would be otherwise incorrectly handled.”

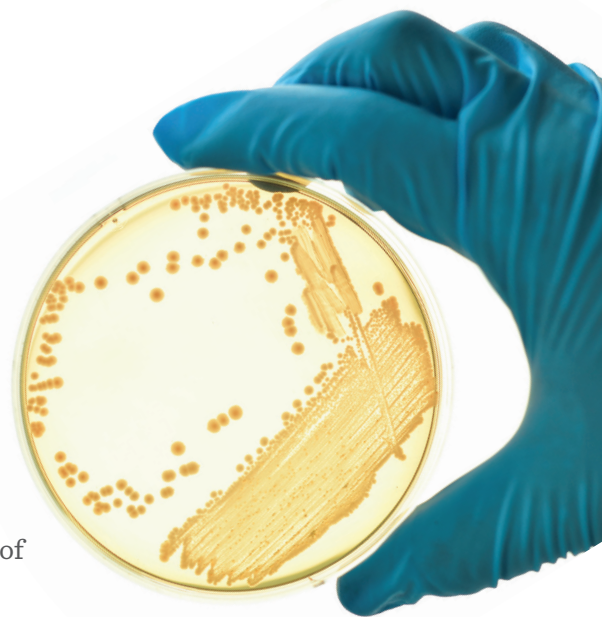
- Elliott Margulies, Illumina

Collaboration in the Cloud

Big data may be a goldmine for discovery, but analysis has become a bottleneck in biomedical and biotechnology research. Companies of all sizes struggle with the challenges of providing the right infrastructure, and EMBL-EBI provides solutions.

One solution is Embassy Cloud, which offers private workspaces within the EMBL-EBI infrastructure so that clients have direct access to data, services and compute alongside their own customised workflows, applications and datasets.

This is a practical, cost-effective alternative to replicating services and downloading vast, public datasets locally. Tenants can access their workspace from anywhere in the world, reducing the need for capital investments in hardware and related operational costs.





Outbreaks: Acting Quickly

At the start of an infectious disease outbreak, public health professionals are increasingly using DNA sequencing to identify small variations between strains of a virus, track transmission and identify the source. This same technology can be used to sequence a whole hospital environment, providing a sensitive tool that enables staff to act quickly to contain a problem at the earliest possible stage.

EMBL-EBI houses crucial information about pathogens and makes it available for retrieval on demand. We also provide a sustainable platform for managing problems with global dimensions.

We are a partner in **COMPARE**, an international collaboration to speed up the detection of and response to infectious disease outbreaks using genome technology. By providing compute and data-sharing infrastructure, we make it possible for clinical, risk-assessment and research professionals to access, analyse and compare crucial information in real time.

“Pathogens don’t respect national borders. The real power of data comes when health professionals and researchers can share and combine it.”
- Paul Kellam, COMPARE project

We are also a partner in the **ZIKAlliance**, a multinational endeavour that investigates clinical, fundamental, environmental and social aspects of ZIKV infection. We help coordinate molecular data for this crucial project, which focuses on pregnant women, environmental factors and epidemiology.

Our teams build the tools collaborators need to reach the best decisions when time is short. Using Embassy Cloud, data resources like the European Nucleotide Archive and advanced tools for analysis, public health projects can find solutions that benefit human and animal health.

Supporting Public Health Communities

COMPARE harmonises the way pathogen surveillance information is shared, providing a new way to combat diseases. COMPARE’s 29 partners include Europe’s leading institutions in the fields of emerging epidemics and food-borne outbreaks, working together to address diseases that can spread from animals and food to humans.

The ZIKAlliance is combatting a global threat that already affects over 70 countries. Its 52 partners combine many disciplines to explore the impact of ZIKV during pregnancy, decipher the natural history of ZIKV and develop a preparedness platform for South America.

Agri-food and Biotechnology

We welcome new collaborations with companies in all sectors of life-science R&D, from large pharma and agri-food industry to biotech and smaller companies. Our goal is to build open, collaborative platforms for translating public research data into new insights, products and solutions.

Securing the Food Supply

Syngenta, a global agri-food company, and the Wellcome Trust supported our expansion of **ChEMBL**, a freely available chemistry data resource, to include information about millions of bioactive molecules from more than 2000 published crop-protection research articles. This resulted in descriptions of over 28 000 new chemicals, their interactions with molecular targets and their biological effects.

Thanks to this work, ChEMBL now offers a rich set of public-domain data on the molecular activity of herbicides, fungicides, and insecticides, integrated with human health information such as toxicity. Researchers developing new crop protection methods can now search for and identify the most promising chemicals and molecular targets easily.

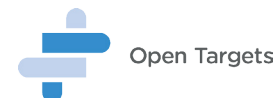
Enzyme Discovery

‘Mining’ the latest scientific knowledge for commercial application demands robust analytical tools. But creating them takes highly specialised knowledge and costly infrastructure, which can be a barrier for product discovery and development. Large companies make use of external companies to provide these services, but such outsourcing is out of reach for most smaller businesses.

An Innovate UK-funded collaboration between EMBL-EBI and Cardiff-based company Biocatalysts Ltd. is facilitating access to high-end analytical tools, allowing companies of all sizes to compete for first advantage and scale up their solutions. **MetXtra**, the project’s novel software platform, enables enzyme discovery through the analysis of large, open-access environmental sample datasets hosted in EMBL-EBI’s Metagenomics service. The platform has enabled the commercialisation of new enzymes, and will boost innovation in the production of fine chemicals, flavours, fragrances, and pharmaceutical products.

- We are a central partner in **TransPLANT**, a European project supporting the analysis of genomic data from crop and model plants.
- **Ensembl Genomes** offers freely available data from important plant, fungal, bacterial and protist species, as well as tools for analyses.
- **EBI Metagenomics** is a platform for analysing and archiving large-scale metagenomic datasets.

On Target for New Medicines



Identifying the right target for a drug is critical to success in drug discovery. Knowing everything about a target right up front—its biology, its function, its genetic drivers—greatly supports this process. When a biological target has a genetic link to a disease, there is a significantly higher chance of success in the clinic, saving years of R&D investment.

Open Targets is transforming the way scientists find, validate and prioritise drug targets. It blends the expertise of world-leading organisations including **GSK**, **EMBL-EBI**, **Biogen** and the **Wellcome Trust Sanger Institute** to identify and validate the causal links between targets, pathways and diseases. Open Targets members combine large-scale experiments (human genome- and pathway-wide cellular, induced-pluripotent stem-cell and organoid screens) with statistical and computational techniques – an approach that helps prioritise biological targets early in drug discovery. This vastly improves the chances of pinpointing the right target and launching a successful medicine.

A Visionary Partnership

Open Targets involves staff from each of its member organisations. Mixed project teams work in a shared space, with joint workshops and events to help staff from different companies make the most of their diverse working practices.

With EMBL-EBI and the Sanger Institute as partners on the Genome Campus, Open Targets is uniquely positioned to generate, integrate and distribute open, high-quality genetic and genomic data on a very large scale.

Enabling Progress

The Open Targets framework applies to any therapeutic area, and currently runs programmes in immunology, oncology and neurodegenerative disease.

Open Targets welcomes new partners who share its vision.

Why Open Targets

- More than 80% of new medicines fail, often in late-stage development, when the cost is highest.
- Establishing genetic links between targets and disease development significantly increases the chances a drug will succeed.
- This demands a combination of skills, expertise, cutting-edge technologies and diverse data types not to be found in a single entity.
- Open Targets brings it together, leveraging the best of its public and private member organisations.

www.opentargets.org

The Industry Programme

An important part of EMBL-EBI's mission is to disseminate cutting-edge technologies to industry. Some 20% of our users are engaged in industrial R&D and our services are constantly evolving to reflect the rapidly changing needs of this crucial sector. As biology becomes more data-driven, the bioindustries are forming pre-competitive collaborations, as well as open-source software and informatics standards to improve efficiency and reduce costs. These are the core elements of our Industry Programme, a forum for interaction and knowledge exchange for those employed at the forefront of industrial bioinformatics.

Industry Programme members

Pharmaceutical



Agri-food, biotech and consumer goods





MANCHESTER
University of Manchester

A14

Cambridge
Research Park

Waterbeach
Station

CAMBRIDGE

NORWICH
Earlham Institute

A14

OXFORD

Wellcome Trust Centre for Human Genetics
Translational Neuroscience & Dementia Research

Cambridge
Science Park

University of Cambridge

School of Clinical Medicine
Autism Research Centre
Alan Turing Institute
Cambridge Cardiovascular
Department of Zoology

- Academic partners
- Other Science parks

Cambridge Biomedical Campus

Cambridge University Hospitals
(Addenbrookes, Papworth, Rosie)
MRC Laboratory of Molecular Biology (LMB)
Cancer Research UK - Cambridge Institute
NIHR Cambridge Biomedical Research Centre
- AstraZeneca
- GlaxoSmithKline Clinical Unit
- e-hospital Project

ARM

Cambridge
Station

M11

Shelford
Station

Babraham Institute

A505

A11

 **WELLCOME
GENOME
CAMPUS**

Base of

 **sanger**

EMBL-EBI

STEVENAGE

GlaxoSmithKline R&D
Bioscience Catalyst

Meldreth
Station

Shepreth
Station

Foxton
Station

Whittlesford
Parkway
Station


Great
Chesterford
Station




LONDON

Francis Crick Institute, Imperial College,
University College London, Wellcome Trust,
LSHTM, Kings College London,
Institute of Cancer research

Chesterford
Research Park
(Illumina)



EMBL Enterprise Management Technology Transfer GmbH (EMBLEM) is an affiliate and the commercial arm of the European Molecular Biology Laboratory (EMBL) and all its outstations. EMBLEM facilitates and accelerates the transfer of innovative technology and know-how from EMBL to industry by way of collaboration, licensing, service provision and creation of spin-out companies.  www.embl-em.de

 www.ebi.ac.uk/about/our-impact
 +44 (0)1223 494 665
 comms@ebi.ac.uk

 @emblebi
 /EMBLEBI
 /EMBLMedia

EMBL-EBI is a part of the European Molecular Biology Laboratory.

European Bioinformatics Institute (EMBL-EBI)
Wellcome Genome Campus
Hinxton, Cambridge CB10 1SD
United Kingdom