

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

(www.ebi.ac.uk/industry 6 +44 (0)1223 494 154

- m +44 (0)1223 494 468
- \sim industry@ebi.ac.uk
- y @emblebi
- f /EMBLEBI You Tube
- EMBLmedia

EMBL-EBI is part of the European Molecular Biology Laboratory (EMBL)

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

You can download this brochure at www.ebi.ac.uk/about/our-impact



The European Bioinformatics Institute · Cambridge Industry Programme



EMBL-EBI and Industry

Our Industry Programme is unique in the world. It is a forum for interaction and knowledge exchange for those working at the forefront of applied bioinformatics, in over 20 major companies with global R&D activities. The programme focuses on precompetitive collaboration, open-source software and informatics standards, which have become essential to improving efficiency and reducing costs for the world's bioindustries.

The European Bioinformatics Institute (EMBL-EBI) is a global leader in the storage, annotation, interrogation and dissemination of large datasets of relevance to the bioindustries. We help companies realise the potential of 'big data' by combining our unique expertise with their own R&D knowledge, significantly enhancing their ability to exploit high-dimensional data to create value for their business.

We see data as a critical tool that can accelerate research and development. Our mission is to provide opportunities for scientists across sectors to make the best possible use of public and proprietary data. This can help companies reduce costs, enhance product selection and validation and streamline their decision-making processes.

Companies with large R&D capacity must ensure high data quality and integrate licensed information with both public and proprietary data. At EMBL-EBI, we help companies build all publicly available data into their local infrastructure so they can add proprietary and licensed information in a secure way.



Going forward, we see our interactions with our industry partners growing stronger, as the flood of data continues to rise. Through our programme and efforts such as the Innovative Medicines Initiative and the Pistoia Alliance, we support precompetitive research collaborations, promote the uptake and utility of open-source software, and steer the development of data standards.

Working closely with industry, we are proud to facilitate the translation of discoveries into new advances that benefit society.

Dr Dominic Clark EMBL-EBI Industry Programme Manager

A unique forum for top companies

Since 1996, our Industry Programme has provided neutral ground where pharmaceutical, agri-food, nutrition and healthcare R&D professionals can discuss shared challenges in bioinformatics.

Precompetitive collaboration is the cornerstone of our Industry Programme. Our interactions with industry have a strong foundation in collaborations with European pharmaceutical R&D, as well as global companies based in the US and Japan. As we expand, we are increasingly well positioned to address the needs of agriculture, nutrition, healthcare, diagnostics and biotechnology research across sectors.

The Industry Programme is subscription-based, with over 20 members representing global companies that make significant use of our data and resources as a core part of their R&D.





Member companies: agri-food and fast-moving consumer goods







Precompetitive collaboration

As biology becomes increasingly data-driven, collaboration in the very early stages of R&D has become crucial for improving efficiency and reducing costs. Together with informatics standards, precompetitive collaboration forms the cornerstone of our Industry Programme.

"We're all facing the same challenges. The Industry Programme workshops are an opportunity for us to collect intelligence about unproven fields in industry, to see what areas and applications we should perhaps focus on. These things are less obvious if you look at the issues only in the context of a single company."

-- Claus Bendtsen, Head of Computational Biology, AstraZeneca

Knowledge exchange	Quarterly meetings
Our Industry Programme members specifically request workshops designed to help them develop data and tools around emerging technologies and shared areas of interest.	We facilitate inter-company interactions in a neutral environment that is exceptionally well suited to discussing bioinformatics challenges collectively faced by our members.
Resolving practical challenges removes roadblocks to discovery, paving the way for new therapeutics, vaccines, consumer goods, agricultural products and novel mechanisms	Open standards
	Our Industry Programme leads the development of open data standards that

Access to experts

Our service teams work closely with researchers in areas of direct relevance to industry, and discuss emerging research areas and opportunities during quarterly meetings and workshops.

that optimise drug-discovery pipelines.

Through our wide network of collaborators, we facilitate access to the world's most pre-eminent researchers in computational biology and related fields. Our Industry Programme leads the development of open data standards that protect our partners from becoming locked into proprietary solutions.

For example, we brought together pharmaceutical companies, public and commercial data providers and academic groups to agree on MIABE: a standard for reporting on bioactive entities. This facilitates the interchange of public data on drug discovery.

Member-driven workshops

Our Industry Programme members request workshops designed to help them develop data and tools around emerging technologies and shared areas of interest. Resolving practical challenges in the uptake of new technologies removes roadblocks to discovery, paving the way to identify new biological targets for treating disease, and novel mechanisms that optimise drug discovery pipelines.

"EMBL-EBI's knowledge-exchange workshops provide an excellent opportunity to engage with stakeholders and key opinion leaders from industry and academic institutes. We were delighted to host the first workshop in the US, which focused on Next Generation Sequencing and emphasised the 1000 Genomes Project and NGS data analysis."

- NR Nirmala, Novartis, Cambridge, USA

Workshops: selected topics

Opportunities in data

- Genome editing in drug discovery and development
- The Human Microbiome: challenges and opportunities for novel therapeutics
- Informatics & -omics for oncology drug resistance
- Bio-pharmaceutical opportunities in proteomics
- Computational systems biology approaches to neurodegeneration drug discovery
- Informatics, genetics and genomics for target selection and validation
- Oryo-electron microscopy/tomography

Data standards

- O Practical use of biomedical ontologies
- Ontologies in agriculture, food and nutrition

- Embedding clinical standards in research: development, implementation and curation
- Expanding IUPAC standards for chemical information: industry applications
- Target tractability assessment

Tools development

- Digital biomarkers: delivering health data for diagnostic medicine
- Single-cell RNA-seq: data generation, processing, resources and analytical tools for applications in cell-based therapies
- O Molecular informatics open-source software
- Success in life-science R&D through User Experience Design
- Antibody developability: aggregation, stability and viscosity
- Deep learning: applications in pharma and agri-food research

Biological Data Resources

The revolution in sequencing technologies has led to dramatic change in the bioindustries. New R&D opportunities arise every day, fed by a wealth of interconnected data that is generated and shared globally. We maintain the world's most comprehensive range of public biological databases, and help scientists everywhere turn information into knowledge.

Molecular archives

European Nucleotide Archive (ENA)

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

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 (EVA)

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

BioSamples

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

Genes, genomes & variation

1000 Genomes

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

EBI Metagenomics

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

Ensembl & Ensembl Genomes

High-quality, integrated annotation on the genomes of vertebrate, plant, fungal, bacterial, protist and metazoan species within an accessible infrastructure, with advanced tools for analysis.

GWAS Catalog

A quality-controlled, manually curated, literaturederived 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

A database 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

Provides functional analysis of proteins by classifying them into families and predicting domains and 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 (PDBe)

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.

EMPIAR

A public archive of raw, 2D electron microscopy images used to build 3D structures.

Chemical biology

ChEMBL

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

ChEBI

Reference chemical structures, nomenclature and ontological classification.

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.

Enzyme Portal

Functional, sequence, nomenclature, substrate, product and cofactor data for enzymes.

Literature

Europe PubMed Central

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

BioStudies

Studies and data that do not fit in traditional archives.

Train online

Free, web-based bioinformatics tutorials and webinars.

7

Innovation and collaboration

Our Industry Programme partners come together to develop standards, data-exchange processes and new resources to support emerging areas of translational research.

Pharmaceutical R&D

EMBL-EBI resources can be leveraged to support decision-making processes at multiple points in the pharma R&D pipeline. The large, public datasets we manage can provide valuable insights that help ease bottlenecks in the R&D process, for example the translation of data about model organisms to information of relevance to clinical research.

We consult regularly with users to ensure our data offerings are aligned to the needs of R&D across sectors and disciplines. Our programme members directly inform the development of public resources and tools, for example the ChEMBL database of bioactive entities and our 'biologically aware' search service.

Although our services have traditionally concentrated on preclinical data, we are increasingly focused on enhanced data integration and the development of custom insight tools – particularly for clinical research data.

Our established pre-competitive collaborations with pharmaceutical companies often involve staff members with experience working in pharmaceutical R&D. Our flexible working culture allows new, multi-disciplinary teams to form easily around individual projects. These teams may comprise visiting staff from companies that participate in our Industry Programme.

Model exchange

To break down barriers to collaboration on pharmacometric modelling, EMBL-EBI developed PharmML, a flexible format for exchanging computational models in pharmaceutical R&D. PharmML is a key component of the IMI-funded DDMoRe model repository, which helps researchers collaborate on models to improve the design of cost-effective, reliable clinical trials of new and repurposed drugs.

Combining data for discovery

Enormous quantities of data about the biological properties of drugs, pesticides, food additives and other molecules are generated in life-science R&D. However, not all useful data are reported, which can waste time, energy and resources. We brought together pharmaceutical companies, public and commercial data providers and academic groups to agree on a standard for describing the effect of a compound on a biological entity. The MIABE standard allows companies to combine data from a large number of discovery programmes, gaining valuable insights into biology.





Innovative medicines

Our Industry Programme serves as an interface between industry-focussed initiatives at EMBL-EBI, the Innovative Medicines Initiative (IMI), the Pistoia Alliance, the Clinical Data Interchange Standards Consortium (CDISC) and many others.

We play a pivotal role in many IMI projects, which remove bottlenecks from the drugdevelopment pipeline. The IMI is a partnership between the European Union, represented by the European Commission, and the European pharmaceutical industry, represented by the European Federation of Pharmaceutical Industries and Associations (EFPIA).

IMI selects and finances basic and translational research projects based on a work programme developed by the European Federation of Pharmaceutical Industries and Associations.

IMI projects at EMBL-EBI

- EBiSC: European Bank for induced pluripotent Stem Cells (iPSC) is a centralised, not-for-profit iPSC bank providing researchers across academia and industry with access to scalable, cost-efficient and consistent, high quality tools for drug discovery.
- EU-AIMS: Academic research, commercial R&D and patient organisations are collaborating to develop translational approaches for autism spectrum disorder treatment.
- EMIF: The European Medical Information Framework is a common framework for patient-level data to facilitate access to diverse medical and research data sources.
- TransQST: The aim of the Translational Quantitative Systems Toxicology project is to improve the understanding of the safety of medicines. The partners are combining existing data and generating new data to support the development of tools for assessing the safety profile of drug candidates before they enter the clinical testing phase.
- Next Generation of Electronic Translational Safety (eTransafe NexGETS): This emerging infrastructure for preclinical and clinical data sharing supports the analysis of animal data for human safety assessment, the discovery of biomarkers and the development of predictive tools for animal and human safety.

Catalysing discovery

EMBL-EBI helps companies focus on innovation by providing access to a sophisticated, large-scale data and compute infrastructure. We work actively with our partners to create data services that support agrochemical, plant breeding and biotechnology R&D.

Plant and animal research

Our data services are used in R&D efforts that result in higher-yield and more stress-resistant crops, and that give insights into health and disease of animals.

- Ensembl Genomes supports research into important pests, pathogens, pollinators, vectors and symbionts, and offers tools for comparative analysis and orthologue prediction.
- PhytoPath provides valuable information on plant pathogenesis at the molecular level.
- The European Variation Archive supports association studies of plants and animals.
- The Ensembl VEP can be used to analyse the functional impact of genetic variants.
- ArrayExpress compares transcriptomics results with public functional genomics data.
- Proprietary chemical information can be linked with bioactive entities and metabolomics data in ChEMBL and MetaboLights.
- Reactome maps genes to biological pathways, and encourages exploration with sophisticated visualisation tools.

Biotechnology

Our bioinformatics resources support research in bioengineering, bioremediation and biocatalysis.

- The European Nucleotide Archive and UniProt – two core EMBL-EBI resources – comprise a molecular catalogue of life. Together, they form a mirror of the natural reservoir of known biological function.
- EBI Metagenomics captures complex new data from under-explored environments, reflecting the rich diversity of life in our land, waters and skies.
- In the EU-funded project Microme, our services and data underpin work in molecular modelling and metabolic engineering.



Open Targets

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

A Visionary Partnership

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 using extensive human genetics and genomics data to transform how drug targets are identified and prioritised. It is a unique partnership, blending the expertise of world-leading organisations including:



Open Targets members combine large-scale experiments with statistical and computational techniques to prioritise biological targets early in drug discovery. This vastly improves the chances of pinpointing the right target and launching a successful medicine. 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.

Mixed project teams with staff from each of the member organisations work in a shared space, with joint workshops and events.

Enabling Progress

The Open Targets R&D framework can be applied to therapeutic areas across the board. Current programmes include oncology, immunology, respiratory and neurodegenerative disease.

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

An open platform

The Open Targets platform, developed at EMBL-EBI, is designed entirely based on userexperience principles. It is a free, open data resource that integrates public-domain data to enable target identification and prioritisation.

New partners

Open Targets is a growing endeavour and welcomes new partners who share its vision and mission to enable development of life-changing medicines in areas of unmet need.

Our missions

Industry

We help disseminate cutting-edge technologies to industry

17

Research

7 1 0 0 1

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

Training

We provide advanced bioinformatics training to scientists at all levels

~)

00

LIII S

Services

We provide data resources free of charge to all facets of the scientific community

ELIXIR

We support, as an ELIXIR node, the coordination of biological data provision throughout Europe

Where to find us

