



EMBL-European Bioinformatics Institute
in a nutshell



Contents

The hub of bioinformatics in Europe	4
Who we are	6
Where we are	7
How we're funded	7
Services	8
Research	10
Training	12
Supporting industry	13
How to find us	14



The hub of bioinformatics in Europe

As we move towards understanding biology at the systems level, access to large data sets holding many different types of biological information has become crucial. However, the high-throughput revolution also provides challenges in coping with the volume of data generated. It is vital that this information is collected, stored and curated in ways that allow its efficient retrieval and exploitation. The European Bioinformatics Institute (EMBL-EBI), which is part of the European Molecular Biology Laboratory (EMBL), is one of the few places in the world that has the resources and expertise to undertake this important task.

The EMBL-EBI grew out of the European Molecular Biology Laboratory's pioneering work to provide public biological databases to the research community. Since being established in Hinxton in 1994, the EMBL-EBI has been a leader in the bioinformatics revolution. We have diversified to provide data resources in all the major molecular domains, expanded to include a broad research base, developed unique ways of supporting our users, and created a user-training programme to enable researchers to make the most of the EBI's data resources and tools.

Strength through collaboration

All our major data resources are the products of international collaborations. We work with other data providers to ensure that our data repositories, and those of our collaborators, are comprehensive and up to date. For example:

- Our nucleotide sequence database, EMBL-Bank, is produced as part of the International Nucleotide Sequence Database Collaboration involving GenBank in the USA and the DNA Databank of Japan.
- The protein sequence and annotation data in the UniProt databases is maintained through the collaborative efforts of three centres, the EBI, the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR) at Georgetown University Medical Center.
- Our macromolecular structure database (MSD) represents the European arm (PDBe) of the worldwide Protein Data Bank (wwPDB), which also includes the Research Collaboratory for Structural Bioinformatics (RCSB) in the United States and the Protein Data Bank of Japan (PDBj). The three organisations act as deposition, data processing and distribution centres for macromolecular structure data.

We also actively participate in international efforts to develop data standards, including the MIAME standard, which specifies the minimum information required to describe a microarray experiment, and the Human Proteome Organisation's Proteomics Standards Initiative (PSI), which has developed data standards for protein-protein interactions, mass spectrometry, gel electrophoresis and general proteomics.

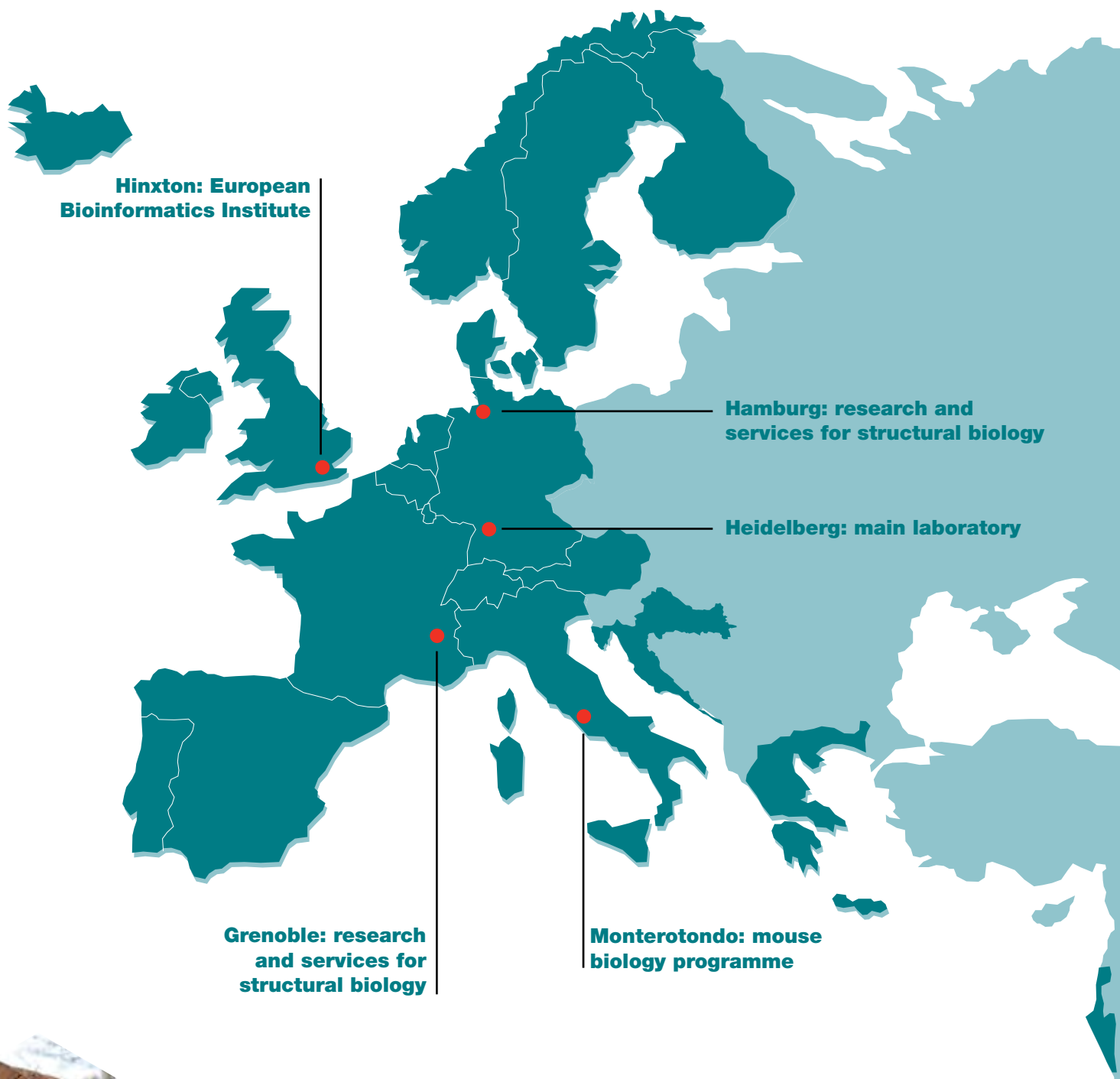
Our mission

- To provide freely available data and bioinformatics services to all facets of the scientific community in ways that promote scientific progress
- To contribute to the advancement of biology through basic investigator-driven research in bioinformatics
- To provide advanced bioinformatics training to scientists at all levels, from PhD students to independent investigators
- To help disseminate cutting-edge technologies to industry

Europe has always been at the forefront of bioinformatics research, but as we move towards the European Union's goal of a single European Research Area, there is a greater need than ever for bioinformatics experts and experimental biologists throughout Europe to work together towards common goals that will expedite biological research. To this aim, the EMBL-EBI coordinates the preparatory phase of the ELIXIR project (www.elixir-europe.org). ELIXIR's objective is to secure funding commitments from government agencies, charities, industry and intergovernmental organisations throughout Europe, to strengthen and sustain a world-class infrastructure for the management and integration of information in the life sciences. ELIXIR is funded from the European Commission's Framework 7 Capacities Programme for Research Infrastructures. The EBI also coordinates three EU-funded networks of excellence and an Integrated Infrastructure Initiative:

- BioSapiens aims to address the current fragmentation of European bioinformatics by creating a virtual research institute for genome annotation and by organising a European school for training in genome annotation.
- EMBRACE will standardise access to bioinformatics resources, enabling data providers to provide well-defined interfaces to their databases that will conform to the same standards. This will allow users to make the most of dispersed data resources.
- ENFIN brings together experimentalists and computational biologists to develop the next generation of informatics resources for systems biology.
- The FELICS Integrated Infrastructure Initiative aims to organise and provide access to a complete range of biomolecular information upon which European life-science research depends.

We are also proactive in organising events that bring computational biologists from all over Europe and beyond together.



A part of the European Molecular Biology Laboratory

The European Molecular Biology Laboratory (EMBL) is a basic research institute funded by public research monies from 20 member states: Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom, and associate member state Australia. Research at EMBL covers the spectrum of molecular biology. The Laboratory has five units: the main Laboratory in Heidelberg, and Outstations in Hinxton (the EMBL-EBI), Grenoble, Hamburg and Monterotondo. The cornerstones of EMBL's mission are: to perform basic, investigator-driven research in molecular biology; to train scientists, students and visitors at all levels; to provide facilities and services for the scientific community; and to actively engage in technology transfer activities.



Who we are

The EMBL-EBI has a dynamic and cosmopolitan atmosphere; our work is highly interdisciplinary and our staff members originate from all over the world. EMBL-EBI's combination of service provision and research offers unique advantages; our commitment to our users ensures that we enjoy access to state-of-the-art computational facilities and systems support in combination with the informal and intellectually stimulating atmosphere of a world-class academic research institute.

Janet Thornton, Director

Janet has been Director of the EMBL-EBI since October 2001. Her research group focuses on understanding biological processes from a structural perspective using computational approaches. After a physics degree, she moved into biophysics at the National Institute for Medical Research, Mill Hill, London. Before moving to the EMBL-EBI she held a joint appointment at University College London and the Bernal Chair in the Crystallography Department at Birkbeck College. She is a Fellow of the Royal Society, a Member of EMBO and a foreign associate of the US National Academy of Sciences.



Graham Cameron, Associate Director

Graham has been Associate Director of the EMBL-EBI since September 2001. He was responsible for the EMBL-EBI's launch in 1993 and ran its information services until 1998, when he became the EMBL-EBI's joint head (with Michael Ashburner). Previous to that he worked at EMBL-Heidelberg in the DNA Data Library, first in software and database design, then leading the project from 1986. His educational background is in psychology, and he has experience in data provision for many disciplines, including primate behaviour, health information, social behaviour and British economic behaviour.



Where we are

The EMBL-EBI shares the Wellcome Trust Genome Campus in Hinxton, near Cambridge (UK), with the Wellcome Trust Sanger Institute and the Wellcome Trust Conference Centre. Set in 55 acres of parkland on the banks of the river Cam, the campus is one of the world's foremost centres for genomics and bioinformatics. We have the advantage of being housed in modern buildings in a beautiful rural setting, whilst being only a few miles away from the academic centre of Cambridge.

The campus provides an exceptionally stimulating environment in which to do top-quality science. It is regularly visited by some of the greatest minds in biomedical research, with workshops and seminars held on campus open to EMBL-EBI staff. We have excellent computational facilities and a well-stocked library. The campus also enjoys a highly active social life. The sports and social club hosts events for all tastes, and there are inter- and intra-institute teams for a wide range of sports, including football, basket-ball and volleyball. We have a gym, a choice of places to eat, and free transport by shuttle bus to Cambridge, Saffron Waldon and some of the surrounding villages.



How we're funded

The EMBL-EBI attracted over €43 million in internal and external funding for 2008. The global importance of our work is reflected in the fact that we attracted over 50% of our funds from external sources, including some beyond Europe. As part of EMBL, the largest part of our funding (generally 40-50%) comes from the governments of EMBL's member states. Other major funders include the European Commission, Wellcome Trust, US National Institutes of Health, UK research councils and our industry partners.

We are grateful for the support of our funders in allowing us to continue and expand our work.

Genomes

Browse and compare vertebrate (and selected invertebrate) genomes in **Ensembl**, which is produced through a collaboration between the EMBL-EBI and the Wellcome Trust Sanger Institute. Completely sequenced genomes, including those of prokaryotes and plants, are available in the various **Ensembl Genomes** interfaces.

DNA and RNA sequence

Find any DNA or RNA sequence in the public domain in **EMBL-Bank**, including completed genomes and sequences associated with patents.

Transcriptomes

Access annotated microarray experiments in **ArrayExpress**, our MIAME-compliant gene expression database.

Protein families, domains and motifs

Perform integrated searches for protein families, motifs and domains with **InterPro**, which combines the strengths of ten member databases to yield a powerful resource for large-scale protein classification and annotation.

Protein-protein interactions

Access information on protein-protein interactions, with supporting evidence, in **IntAct**. Build graphical views of interaction networks and annotate them with functional information. Data in IntAct conform to the **HUPO Proteomics Standards Initiative's Molecular Interaction format**.

Services

The EMBL-EBI maintains the world's most comprehensive range of molecular databases. We are the European node for globally coordinated efforts to collect and disseminate biological data. Many of our databases are household names to biologists – they include EMBL-Bank (DNA and RNA sequences), Ensembl (genomes), ArrayExpress (microarray data), UniProt (protein sequences), MSD (macromolecular structures), IntAct (protein-protein interactions), Reactome (pathways) and CiteXplore (our portal to the scientific literature). The details of each database vary, but they all uphold the same principles of service provision [see box].

Our data resources are free

We are the custodians (not the owners) of biological data provided by the community, and progress in biological research depends on unrestricted open access to these data. Our website (www.ebi.ac.uk) provides direct access to all our data resources and tools, from quick links on the main page or from pull-down menus. This is complemented by our search engine, the EB-eye, an integrated search interface that allows you to search all of our databases at once. You can refine your search, confine the search to specific data resources, and move seamlessly from one data resource to another. We provide tools that allow you to submit your own data, perform complex queries across multiple databases, analyse data and view the results in different ways. You can also download data and software from ftp.ebi.ac.uk, and access a growing number of our resources programmatically using web services.

Supporting our users

We want to make sure that all our users, who are responsible for over 2.5 million requests on our website each day, have all the information they need to make the most of our resources. If you need some guidance on how to make the most of a service, we provide help pages at www.ebi.ac.uk/help as well as the 2Can user support portal at www.ebi.ac.uk/2can. You can also e-mail our helpdesk at support@ebi.ac.uk and we'll get back to you within two working days.

Principles of service provision

Accessibility - Our data and tools are freely available to the research community, without restriction (except in the case of personal genomics information where personal information is withheld to protect anonymity).

Comprehensive data sets - Where several publicly available repositories exist, we have negotiated data-sharing agreements to ensure that our resources contain comprehensive and up-to-date data sets. We also negotiate with publishers to ensure that, wherever practicable, biological data are placed in a

Literature

Search the scientific literature and automatically add links to biological data resources using [CiteXplore](#).

Protein sequence

Use [UniProt](#) – the world’s most comprehensive catalogue of information on proteins – to gain user-friendly access to all the information previously in Swiss-Prot, TrEMBL and PIR-PSD.

Protein structure

Find any of the Protein Data Bank’s structures in the [Macromolecular Structure Database](#). View predicted quaternary structures, interactions with ligands and structure comparisons.

Small molecules

Bridge the gap between the protein world and that of small molecules with [ChEBI](#); explore structure–activity relationships of drugs and drug-like molecules with [ChEMBL](#).

Pathways

Navigate a map of human biological pathways, ranging from metabolic processes to hormonal signalling, in [Reactome](#).

public repository as part of the publication process and cross-referenced in the relevant publication.

Quality - Our databases are enhanced through annotation: features of the genes or proteins stored in them are extracted from other sources, defined and interpreted. Much of our annotation is performed by highly qualified biologists, and the automated annotation that we do is subjected to rigorous quality control.

Compatibility - EMBL-EBI has possibly done more than any other organisation in the world to promote the adoption of standards in bioinformatics; the development of these standards promotes data sharing.

Portability - All our data sets are available for download from the EMBL-EBI website. In many cases the entire software system can be downloaded and installed locally.

Research

The EMBL-EBI provides a unique environment for bioinformatics research. We have a broad palette of research interests that complement our data resources, and these two strands of activity are mutually supportive. Although scientists worldwide have access to our data resources, being part of Europe's main centre for biological data provision offers clear advantages for bioinformatics research. EMBL-EBI researchers have the expertise of hundreds of bioinformaticians close at hand – even in the lunch queue or on the volleyball pitch. Our research teams aim to understand biology through the development of new approaches to interpreting biological data. The services teams, who develop and maintain our data resources, also perform research to create new services and enhance our existing ones. Here we summarise the EBI's research activities performed in both the research and services groups.

Rolf Apweiler: **Protein data resources**

Rolf's team is responsible for the protein and proteomics-related activities of the PANDA group, which include the production of protein sequence and protein family databases at the EMBL-EBI (such as UniProt). The group's research activities involve developing methods to calculate similarity scores for comparison of experiments on a large scale, integrated analysis of expression quantitative trait loci to study mechanisms underlying natural variation of gene expression and improving the automatic annotation of UniProtKB/TrEMBL and integration with other resources.

Paul Bertone: **Differentiation and development**

Paul's group investigates the cellular and molecular processes underlying mammalian stem cell differentiation. Using a combination of experimental and computational approaches, they are working towards a better understanding of which factors confer the capacity for self-renewal, and the events that trigger the onset of lineage commitment. The generation of large-scale data from functional genomic and proteomic experiments will define the global regulatory influence of key transcription factors, signalling genes and non-coding RNAs involved in controlling vertebrate embryogenesis.



Ewan Birney: **Nucleotide data resources**

Ewan's team is responsible for the nucleotide sequence resources of the PANDA group, which includes EMBL-Bank and the Ensembl genome browser. The group's research activities are synergistic with the nucleotide service projects and focus on algorithmic methods for genome analysis, including de Bruijn graph representation of DNA sequence, the evolution of promoter regions, weak binding motifs and the broader scale evolution of regulatory regions and genetic epistasis models.

Alvis Brazma: **Microarray informatics**

Alvis's team is responsible for the development of the ArrayExpress resource for gene expression and related functional genomics data. In addition, the group's research focuses on the integration and meta-analysis of different data sets, including genomics and medical data, with the goal of understanding and modelling basic biological processes or diagnostics. Other research activities include high-throughput data analysis and development of algorithms for systems biology and biomedical informatics research.

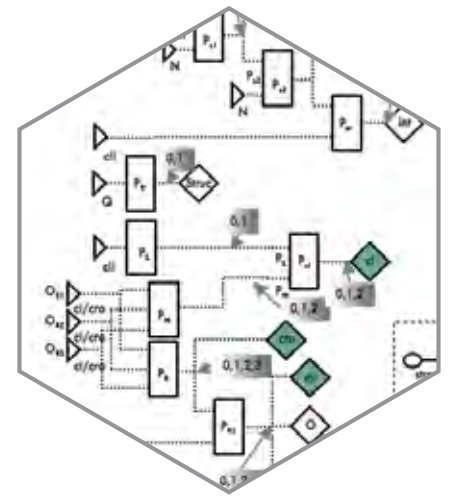
Anton Enright: **RNA genomics**

Anton's group has a particular interest in the function of small non-coding RNAs including microRNAs (miRNAs) and piwi-interacting RNAs (piRNAs). The lab develops computational tools, systems and algorithms for functional genomics of small RNAs. The group collaborates extensively with experimental labs in order to build better data sets for helping us to understand these important regulatory molecules. The group also has interests in the visualisation of large data sets, graph-clustering and analysis of biological networks.



Paul Flicek:
Vertebrate genomics

Paul's team is a combined service and research group that brings together the tools and resources of the Ensembl genome analysis framework for comparative genomics, variation and functional genomics. The team also develops the large scale bioinformatics infrastructure for human variation projects such as the 1000 Genomes Project and European Genotype Archive as a first step towards annotating all human variation. The team's research seeks to create models which integrate diverse data types such as extensive comparative sequencing, DNA-protein interactions, epigenetic modifications, and the DNA sequence itself in the search for every important piece of the genome.



Nick Goldman:
Evolutionary analysis of sequence data

Nick's group is developing better methods for analysing DNA and amino acid sequences to study evolution, and uses information from evolutionary studies to improve other data analyses in bioinformatics. Their work includes the theoretical basis of phylogenetic analyses of sequence data. They also focus on devising better mathematical models to describe sequence evolution. By seeing which biological features add most to the accuracy of the models, they hope to find out which proposed evolutionary forces are responsible for the patterns of sequence variation within and among species.

Wolfgang Huber:
Functional genomics

Wolfgang's group is developing advanced mathematical and statistical methods for understanding functional genomics data. This includes: analysis methods for microarrays; next-generation sequencing and other new functional genomic technologies; automated phenotyping in genetic screens; and modelling of genetic interaction networks. The development of scientific software complements the group's research, and they are a core contributor to the Bioconductor project, an open source and open development software project for the analysis and comprehension of genomic data.

Nicolas Le Novère:
Computational systems biology

Nicolas's group aims to gain a better understanding of signal transduction in neurons by building computational models. A strong focus is the molecular and cellular basis of neuroadaptation in neurons of the basal ganglia. Modelling requires various types of software, including design environments and simulators. Nicolas has been instrumental in the development of Systems Biology Markup Language (SBML), which is designed to facilitate the exchange of biological models between different types of software. His group also hosts the BioModels database, a repository for annotated biological models.

Nick Luscombe:
Genome-scale analysis of regulatory systems

Nick's group uses computational techniques to analyse transcriptional regulatory systems on a genomic scale. Much of their work focuses on the regulatory system in yeast. By integrating diverse information – from its genome sequence to the results of functional genomics experiments – they can study the regulatory system of an entire organism. They have also expanded their interests to understanding regulation in enterobacteria and humans.

Dietrich Rebholz-Schuhmann:
Extraction of facts from the scientific literature

Dietrich's group performs research on text-mining methods to automatically extract information from the scientific literature and align the results with the content of the EMBL-EBI's databases. The best performing technical solutions are made freely available as research tools, such as Whatizit, a modular text-mining infrastructure. The group also works towards standardisation of the literature and better support for reviewers, authors and researchers who want to automatically exploit literature (UKPMC, CALBC). Collaborations with ChEBI, ArrayExpress and other groups contribute to this goal.

Janet Thornton:
Proteins – structure, function and evolution

Janet's research group aims to understand more about how biology works at the molecular level. They analyse how enzymes perform catalysis, how these molecules recognise their cognate ligands, and how proteins and organisms have evolved to create life. They develop and use novel computational methods, gathering data either from the literature or by mining the data resources, to answer specific questions. Much of their research is collaborative, involving either experimentalists or other computational biologists.

Training

Training future bioinformaticians

Bioinformatics is a rapidly expanding discipline that touches on all areas of biology – pure and applied. There is a global shortage of trained bioinformaticians, and the EMBL-EBI fulfils an important function in working towards reducing this shortage.

Graduate training: We have a lively graduate community that is part of EMBL's International PhD Programme. Students can graduate from the university of their choice. Many of the EMBL-EBI's PhD students join the University of Cambridge; they participate fully in Cambridge graduate life and leave the EMBL-EBI with a PhD from the University of Cambridge. We encourage potential candidates to visit www.ebi.ac.uk/training/studentships/ for more information.

Postdocs: Postdoctoral fellowships are available in our research groups. We are classed as an 'academic analogue' in the UK, and can apply to the Wellcome Trust and the UK research councils for funding for postdoctoral fellowships. We encourage potential candidates to contact the group leader of their choice to discuss the possibility of doing postdoctoral research at the EMBL-EBI.

Visitors: There are opportunities for researchers at all stages of their career to spend time at the EBI. See www.ebi.ac.uk/training/Visitors_Programme/ for more information.



Training days: We hold training days for masters and PhD students from other institutions. Information about our training days and other EMBL-EBI events is available at www.ebi.ac.uk/training or can be obtained by contacting training@ebi.ac.uk.

User training

Our user-training programme aims to equip biomedical researchers with the skills to make the most of their data. We have a three-pronged approach to user training:

Hands-on training at the EMBL-EBI

This programme, held in our purpose-built IT training suite, is especially designed to give experimental biologists expert tuition that will help them to get to grips with their data using the EBI's data resources. Our courses, which are typically 2-5 days long, feature: a combination of hands-on and lecture-based tuition; direct contact with expert trainers; high-quality training materials; and a well-equipped training room. See the full programme at www.ebi.ac.uk/training/handson.

The Bioinformatics Roadshow

As part of the EU-funded FELICS Integrated Infrastructure Initiative, the EBI coordinates the Bioinformatics Roadshow, in collaboration with: the Swiss Institute of Bioinformatics (SIB), the European Patent Office and the BRENDA project. The roadshow provides a mobile training programme covering Europe's most widely used biological data resources. Roadshows are typically held over two days (this is decided by the host) and are modular in nature. The roadshow 'modules' can be mixed and matched to create bespoke courses suitable for the host institute. Learn more about how to host a roadshow at www.ebi.ac.uk/training/roadshow.

eLearning

We also recognise the time and financial restrictions faced by researchers and are developing an eLearning portal to support and complement our face-to-face training programmes. Several courses are already available and you can sign up at www.ebi.ac.uk/training/elearning.



Supporting industry

A comprehensive programme for big business

Advances in bioinformatics are having a major impact on industry. We have been supporting the needs of industry for over 12 years through our Industry Programme, which: kick-starts research of importance to industry; provides expert training; develops bioinformatics standards; helps its partners with technical development; and provides regular networking opportunities. The programme is driven by its partners and membership is by invitation. Our industry partners include large multinational companies in the biotech, pharmaceutical, agricultural, nutrition, personal-care and medical devices industries. The programme is funded by annual member subscriptions.



Industry Partners

AstraZeneca
Bayer Schering Pharma
Boehringer Ingelheim
Eli Lilly and Company
Galderma
GlaxoSmithKline
F. Hoffmann-La Roche
Johnson & Johnson Pharmaceutical Research & Development
Merck Serono S.A.
Nestlé Research Centre
Philips Research
Pfizer Ltd
Sanofi-Aventis
Syngenta
Unilever

Support for small companies

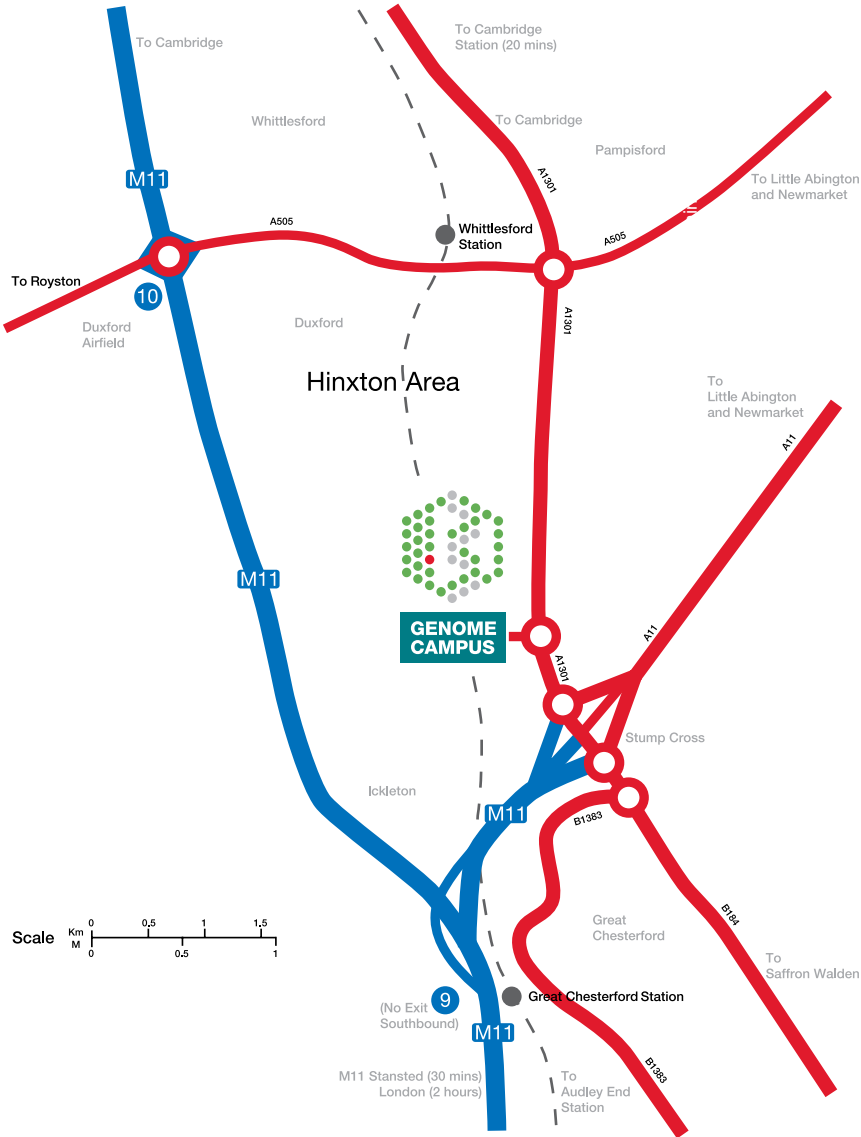
In addition to supporting large companies, the EMBL-EBI also seeks to serve small-to-medium enterprises (SMEs). The SME Support Forum aims to provide beneficial support at minimal cost to participants. The Support Forum meets on an annual basis, offering discussion and networking opportunities. Forum members also have access to: expert tuition; opportunities for technical development through collaboration; and consultancy on an ad hoc basis. Members of the SME Support Forum range from drug discovery and biotech start-ups to bioinformatics service providers.

For more information on how your company can benefit from involvement in the Industry Programme, visit the Industry Support pages at www.ebi.ac.uk/industry/. To learn more about joining the programme, please contact Dominic Clark, our Industry Programme Coordinator clark@ebi.ac.uk.

How to find us



EMBL-European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge
CB10 1SD, UK



Copyright © EMBL-EBI 2008.

Photos courtesy of EMBL Photolab and Wellcome Images.

Designed and produced by PickeringHutchins Ltd www.pickeringhutchins.com

EMBL-EBI



EMBL-European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge
CB10 1SD, UK

Tel: +44 (0) 1223 494444
Fax: +44 (0) 1223 494468
e-mail: support@ebi.ac.uk
web: www.ebi.ac.uk

A part of the European Molecular Biology Laboratory