

EMBL-European Bioinformatics Institute

EMBL-EBI
Institute for
Data Science

EMBL-EBI

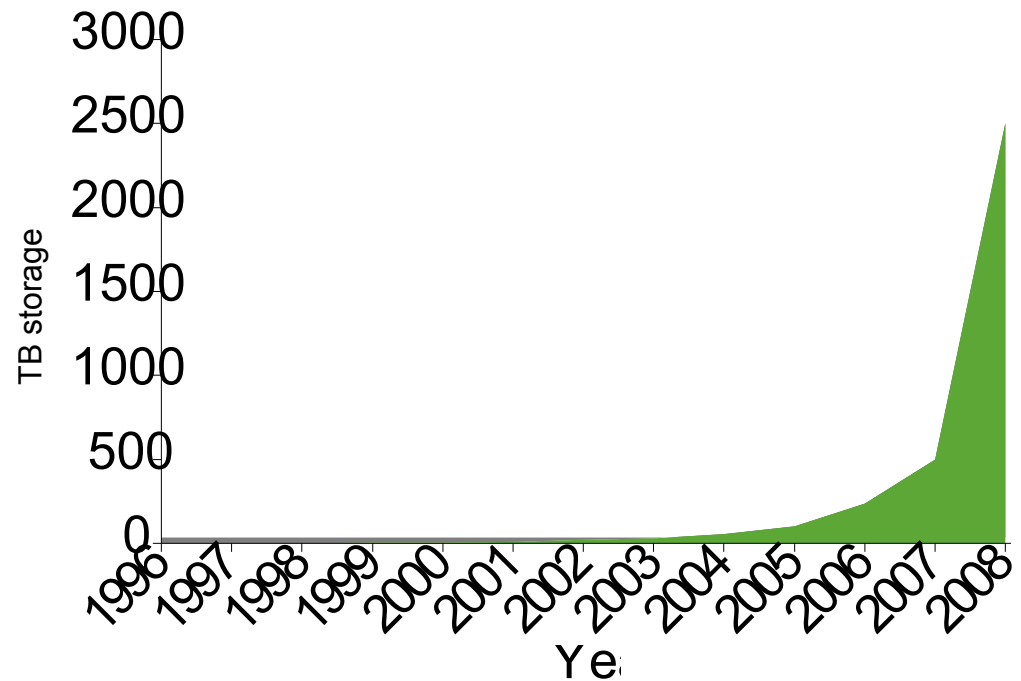


What is bioinformatics?

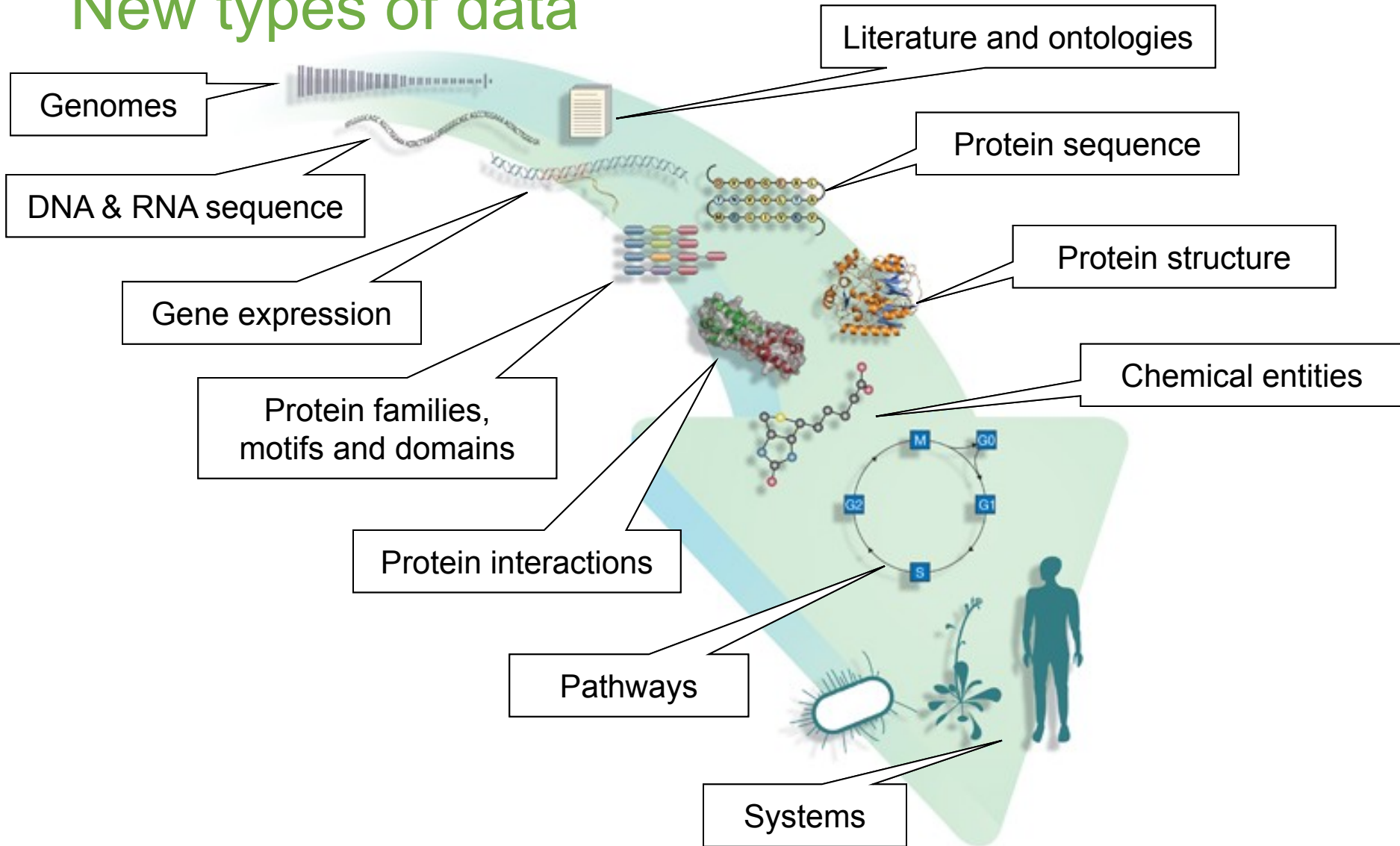
- The science of storing, retrieving and analysing large amounts of biological information
- An interdisciplinary science, involving biologists, computer scientists and mathematicians
- At the heart of modern biology

Biology is changing

- Data explosion and new types of data
- High-throughput biology
- Emphasis on systems, not reductionism
- Growth of applied biology – molecular medicine, agriculture, food, environmental sciences...



New types of data



What is EMBL-EBI?

Based on the Wellcome
Trust Genome Campus
near Cambridge, UK

Part of the **European
Molecular Biology
Laboratory**

Non-profit organisation



The five branches of EMBL

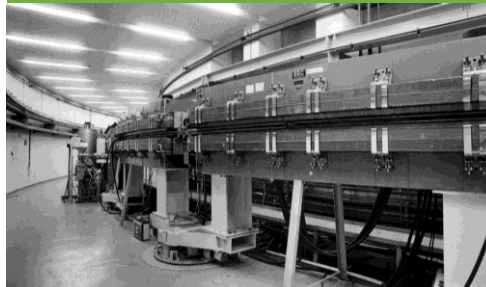
- EMBL is a basic research institute funded by public research monies from 20 member states.
- 1400 staff, over 60 nationalities.

Heidelberg



Basic research in
molecular biology
Administration
EMBO

Hamburg



Structural biology

Hinxton



Bioinformatics

Grenoble



Structural biology

Monterotondo

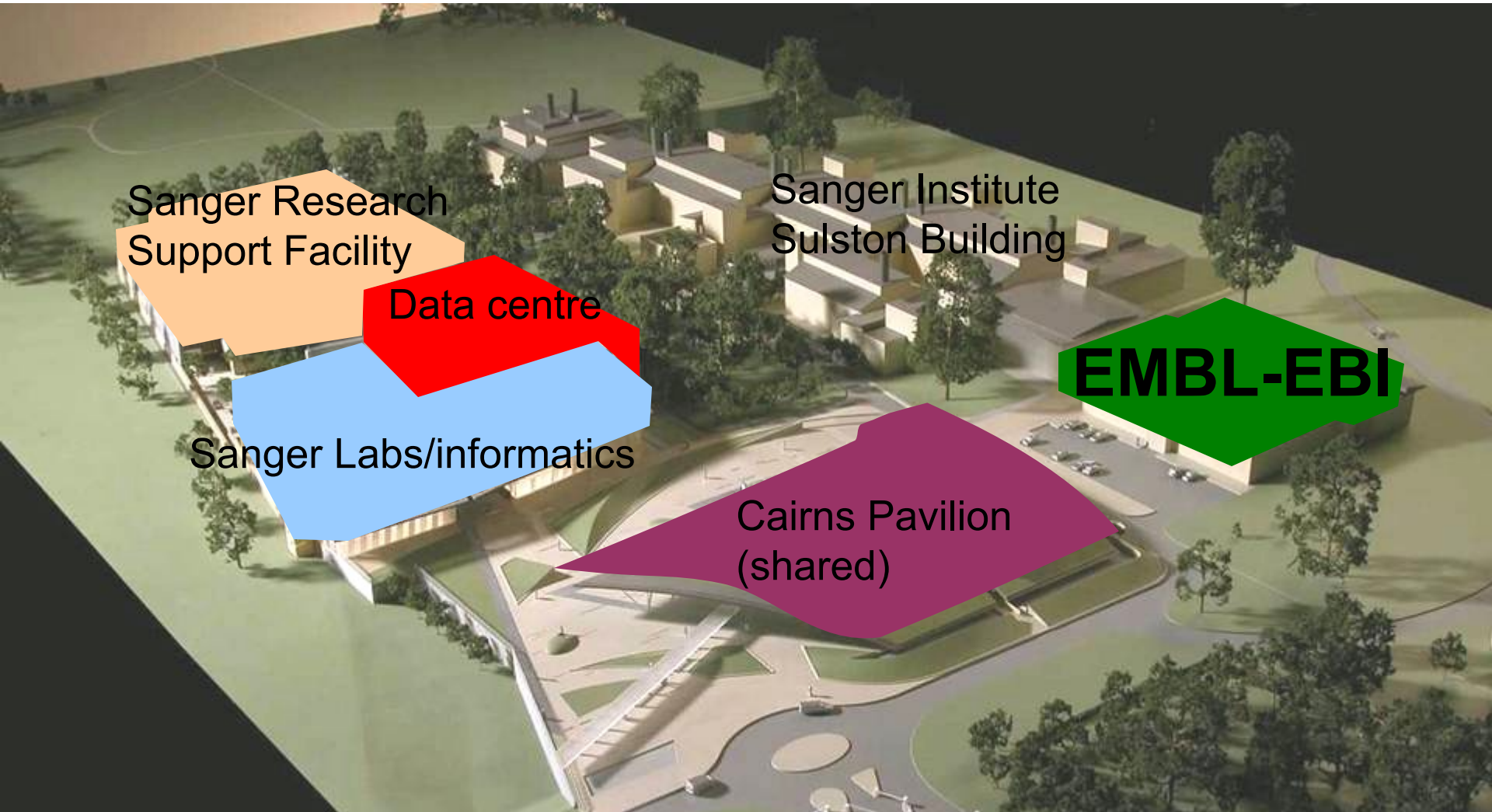


Mouse biology

EMBL-EBI's mission

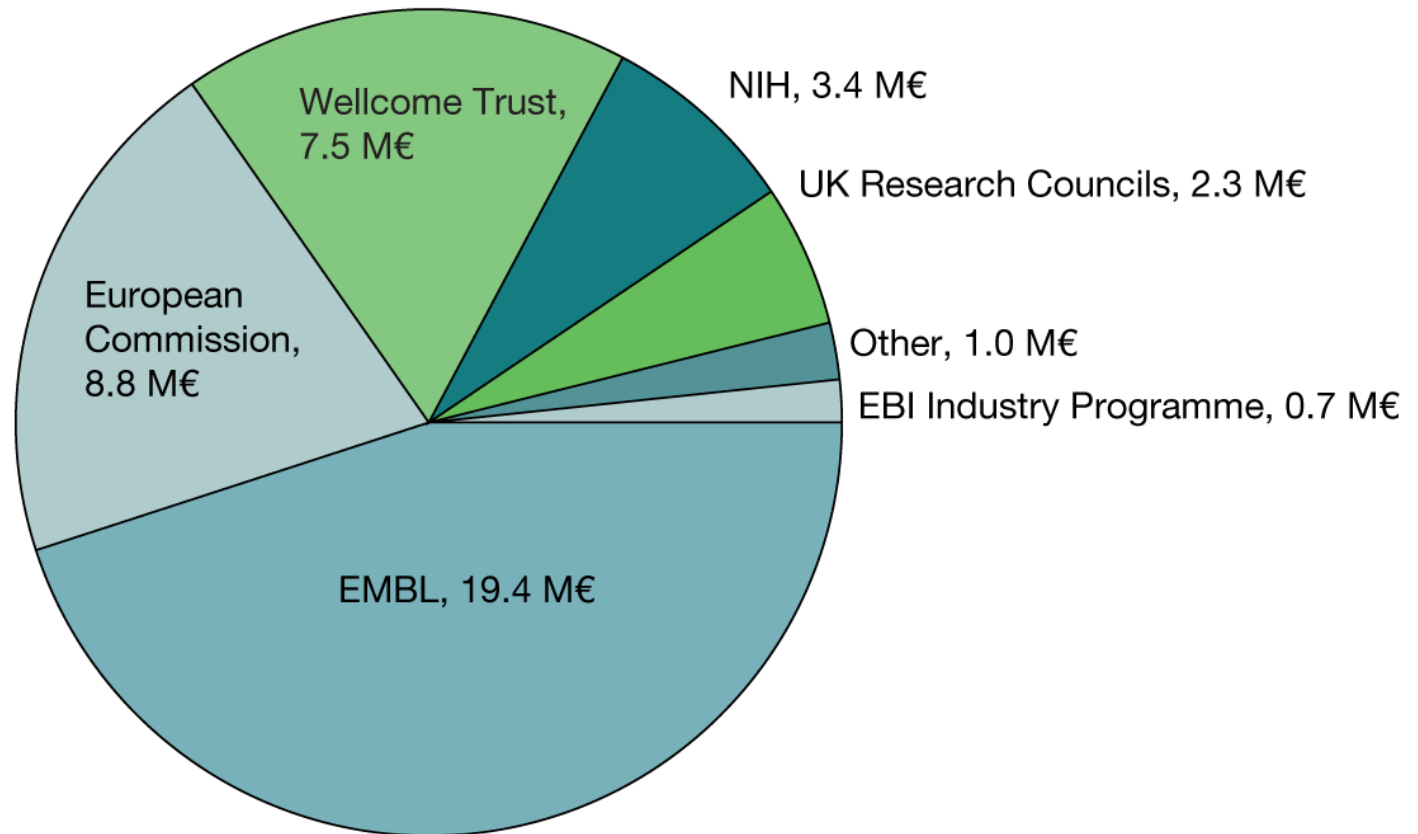
- To provide **freely available data and bioinformatics services** to all facets of the scientific community in ways that promote scientific progress
- To contribute to the advancement of biology through basic investigator-driven **research** in bioinformatics
- To provide advanced bioinformatics **training**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

The Wellcome Trust Genome Campus

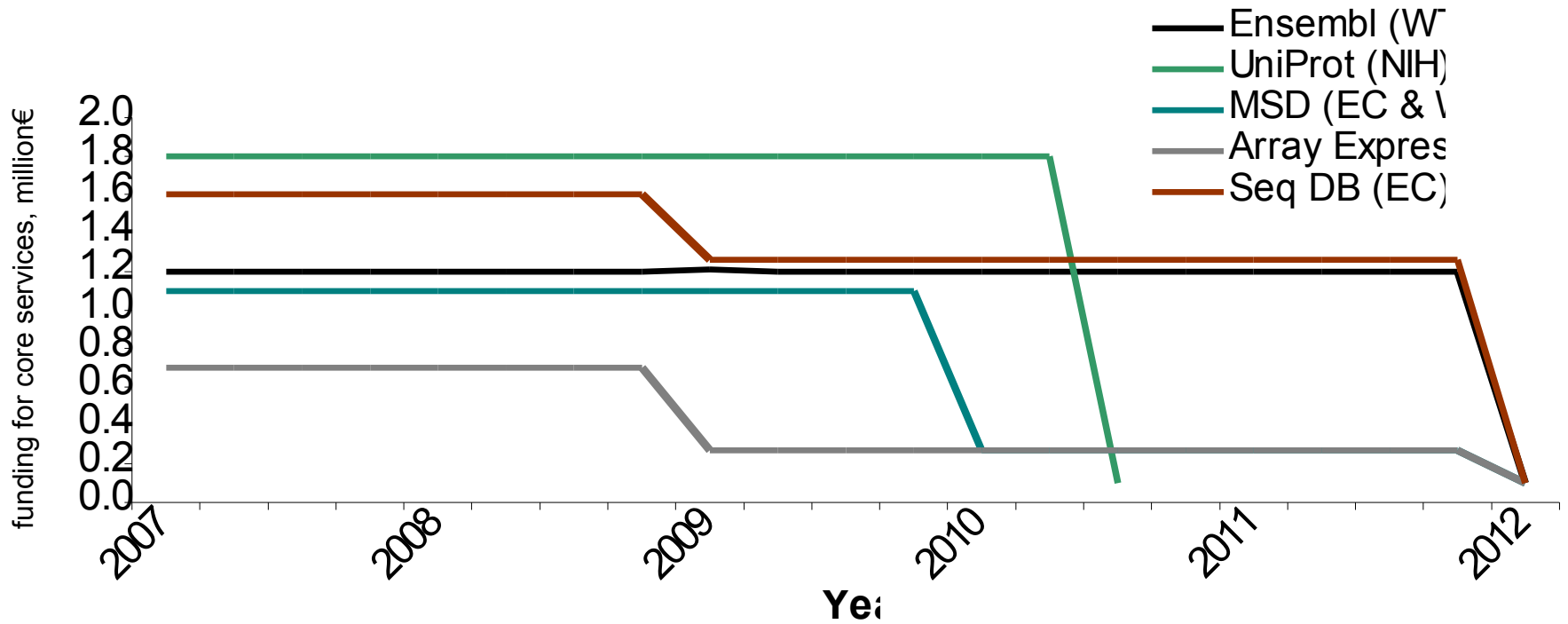


EMBL-EBI funding

Sources of funding for the year as of November 2008. The Wellcome Trust also supports us through provision of our buildings.



but...





- The preparatory phase of ELIXIR, an EU-funded project to agree upon the future bioinformatics infrastructure for Europe, began on 1 November 2007
- **Anyone** involved with bioinformatics in Europe is a stakeholder in this process
- **Outcome** - the resulting memorandum of understanding among EU member states will pave the way towards a more stable footing for Europe's core data resources in the future
- Next **stakeholder meeting** will be held in Copenhagen, Denmark on 19–20 May 2009

See www.elixir-europe.org/ to register



Services

www.ebi.ac.uk/services

EMBL-EBI



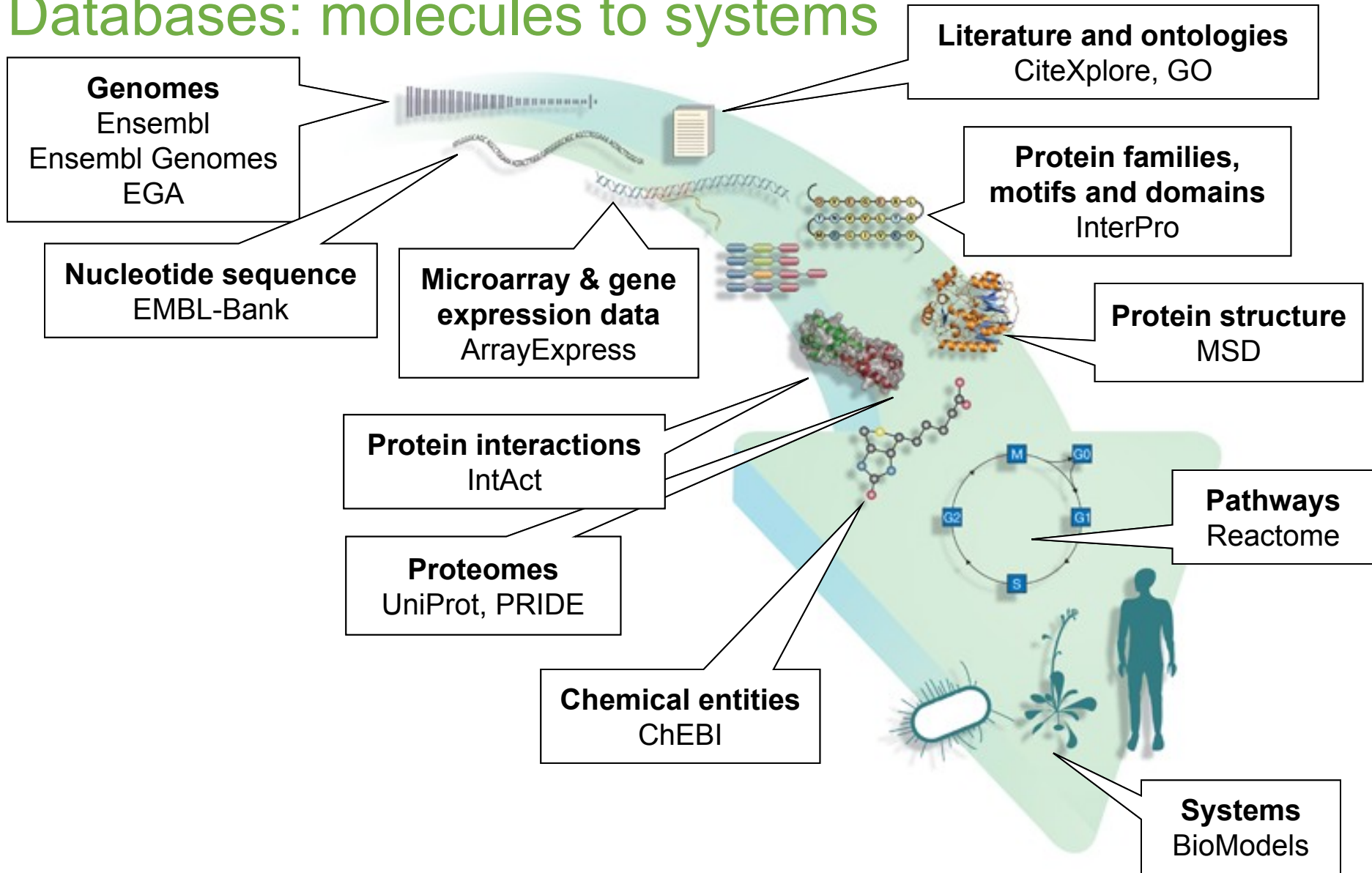
Key facts about services

- **European node** for globally coordinated data collection and dissemination projects
- Core databases produced in **collaboration** with other world leaders, including NCBI (US), National Institute of Genetics (Japan), Swiss Institute of Bioinformatics, Cold Spring Harbor Laboratory (US)
- The world's most **comprehensive** collection of molecular databases

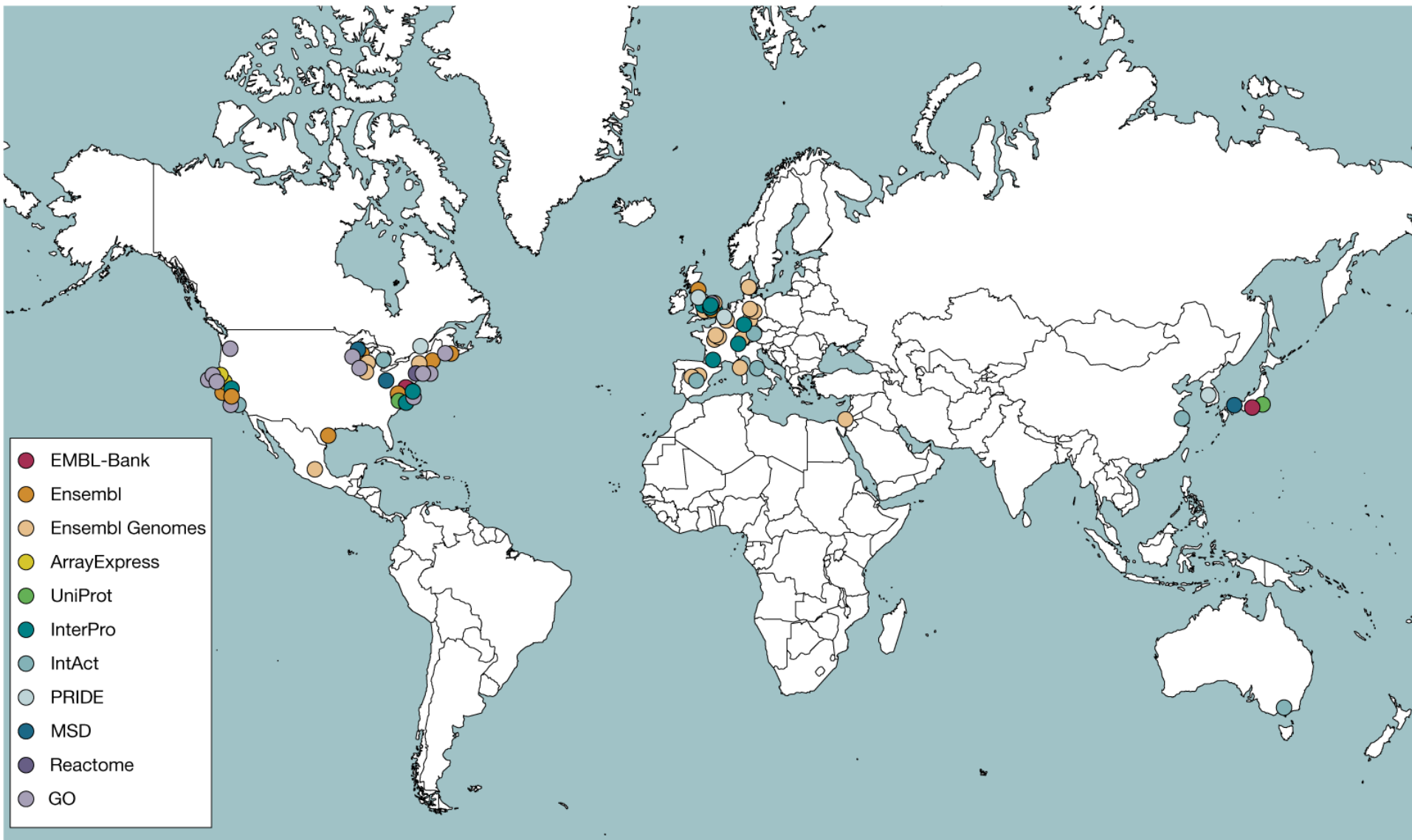
Principles of service provision

- **Accessibility** – all data and tools freely available without restriction
- **Compatibility** – we develop and promote the use of standards in bioinformatics
- **Comprehensive data sets** – agreements with other data providers ensure that our resources contain comprehensive and up-to-date data; agreements with publishers ensure that published data are placed in a public repository at the earliest opportunity
- **Portability** – data and software can be downloaded and installed locally
- **Quality** – Our databases are enhanced through annotation and cross-referencing

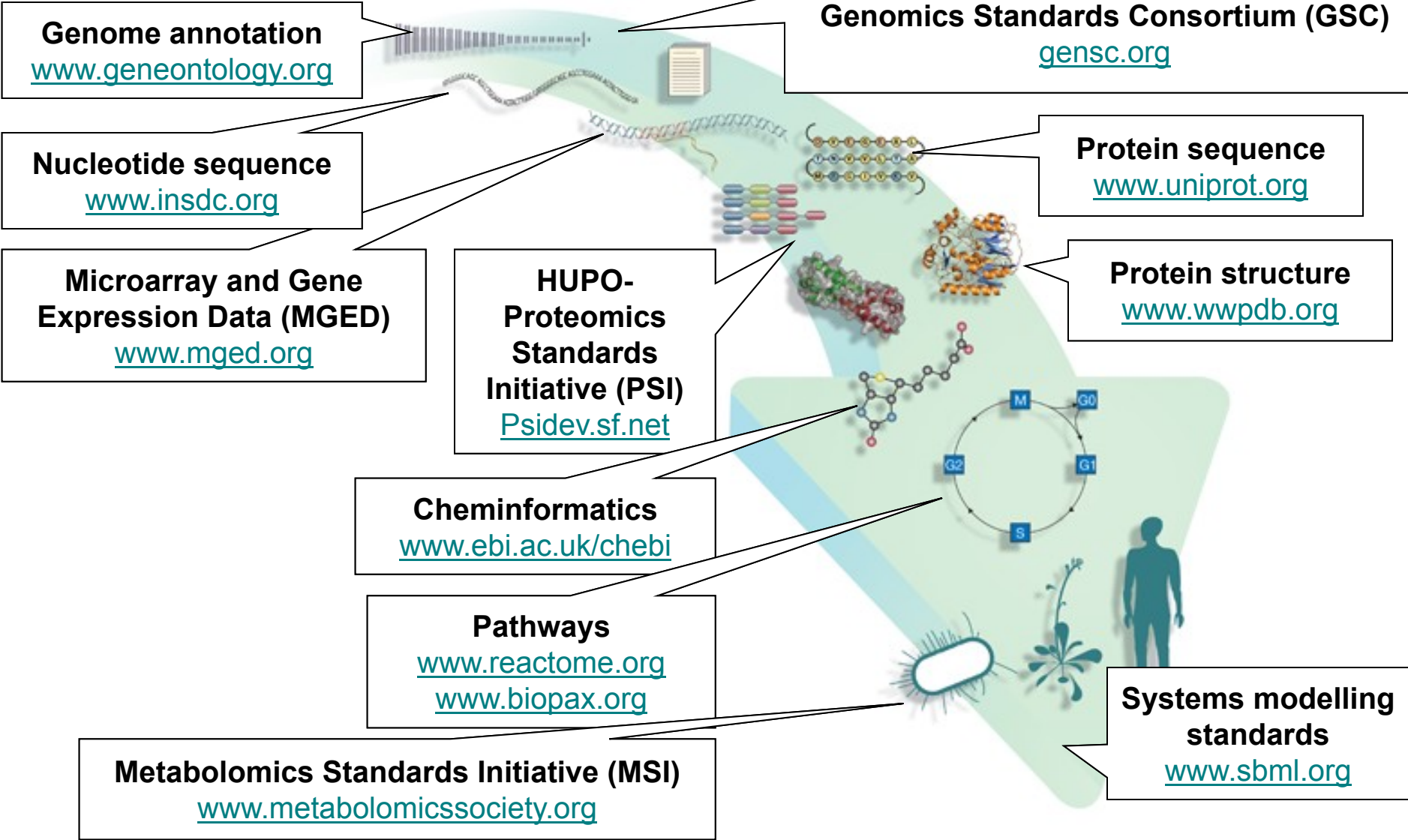
Databases: molecules to systems



Database collaborations



Standards development – international collaborations



EBI website and search engine EB-eye

Search all main databases in one go

Domain	Count
Genomes	7
Ensembl	7
Nucleotide Sequences	15
EMBL	14
EMBL (Update)	1
EMBL (Whole Genome Shotgun)	0
Protein Sequences	2
UniProt	2
Macromolecular Structures	11
MSD/PDB	11
Small molecules	0
Reactions & Pathways	0
Reactome	0
Protein Families	6
InterPro	6
Enzymes	0
Intenz	0
Literature	875
Medline	841
Patents	34
Ontologies	0
GO	0
EBI Web Site	2
2can Support Portal	0
EBI Members and Groups	0
Main sections	2

Advanced Search - Domain selection

Please select the domain you want to search.

The list on the left side of this panel represents all the domains which can be searched. They are organized as a hierarchy and can only be selected individually.

- The domains in bold are categories. Searching a category will result in searching all the fields in all the data sources belonging to this category.
- The other domains are data sources. When selecting a data source, it's possible to narrow the search to particular fields of this data source.

Advanced search: drill down to specific fields in specific databases

Refine your search:

Search for *brca1 human in All the EBI*

with the following keywords

Refine

Refine your search

Genomes 2: Ensembl Genomes



Ensembl Metazoa

Ensembl-like genome browser for non-vertebrate species



Using view options, you can select to view only the current gene or the entire expanded gene tree.

Ensembl Bacteria



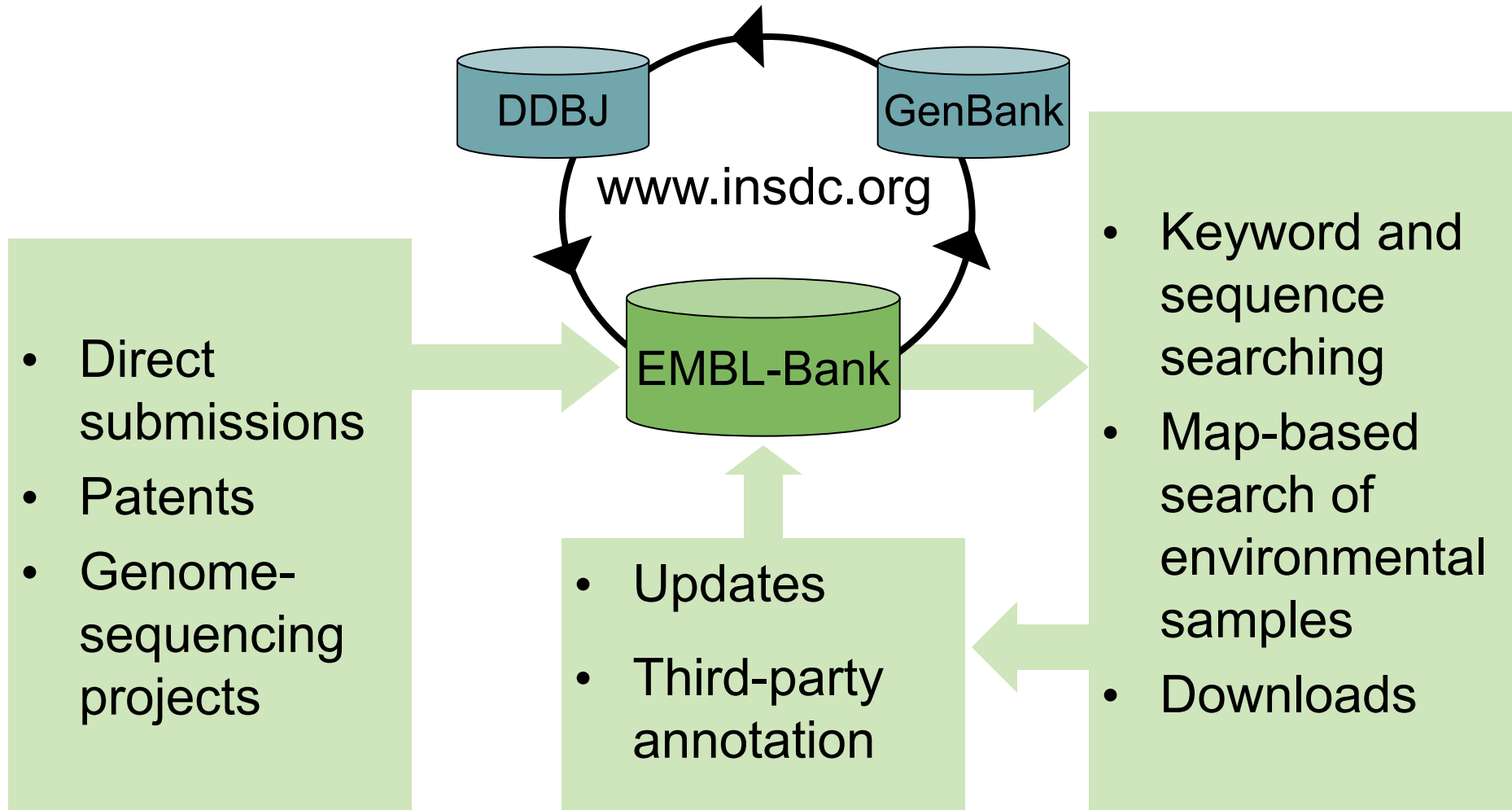
Select Orthologue view to see putative orthologues.



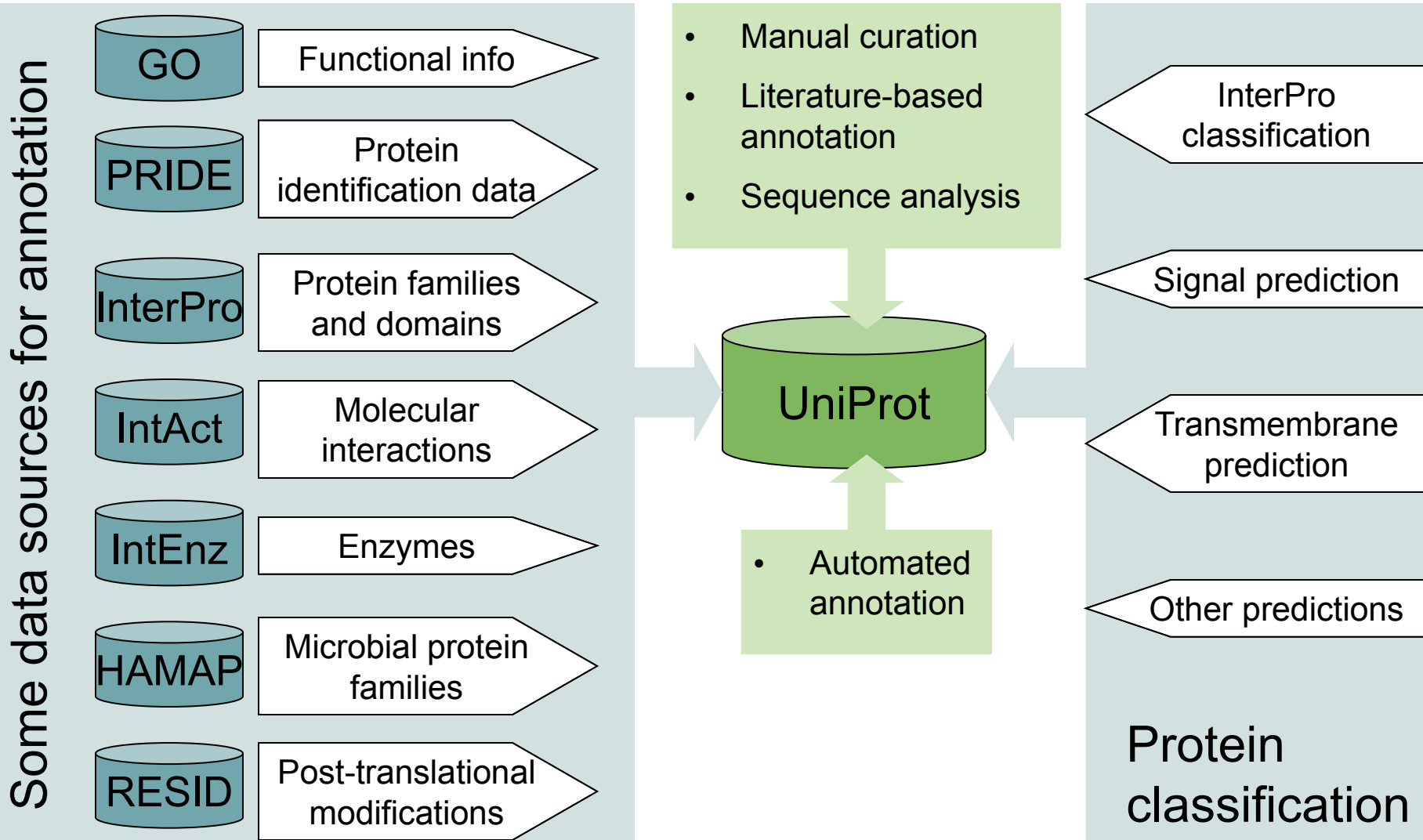
Across species

View options

Nucleotides: EMBL-Bank



Protein sequence: UniProt

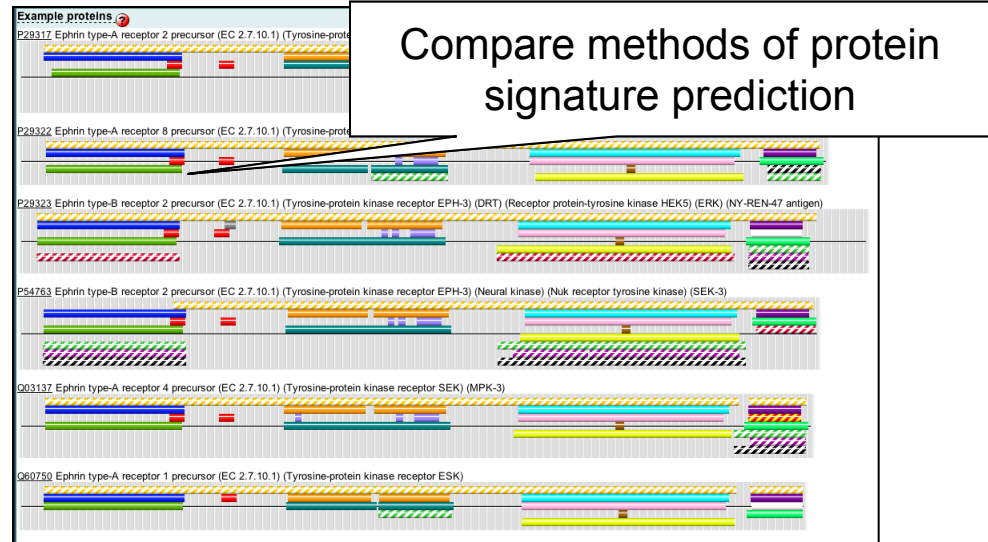


Protein families, motifs and domains: InterPro

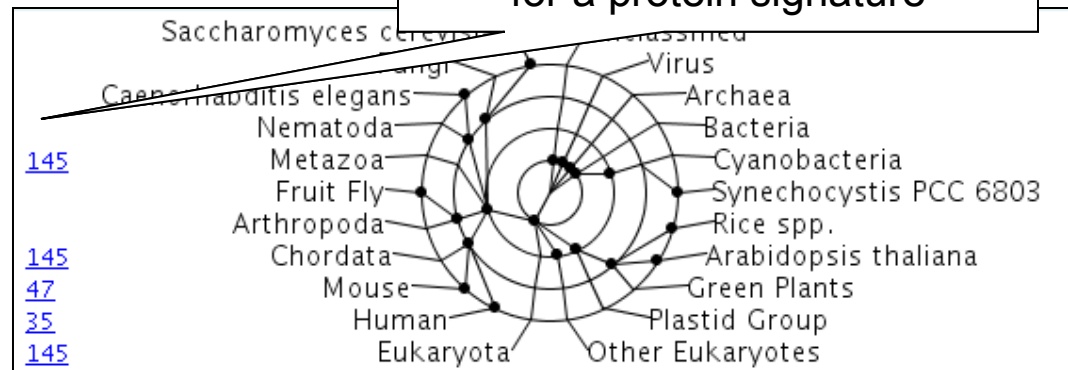
Powerful tool for protein classification, integrating several methods into one resource

Count	Example Code	Architecture
52	Q09127 IDA1090.3961x2.1245[719].1660	Eph-rcpt lig (red) FN III (blue) Tyr kinase (green) Prot kinase (orange) SAM (teal)
32	Q42422 IDA1090.3961x2.719[1245].1660	Eph-rcpt lig (red) FN III (blue) Prot kinase (orange) Tyr kinase (green) SAM (teal)
16	Q08644 IDA1090.3961x2.719[1245].1660[11510]	Eph-rcpt lig (red) FN III (blue) Prot kinase (orange) Tyr kinase (green) SAM (teal) SAM 2 (yellow)
15	Q3UPV5 IDA1090	Eph-rcpt lig (red)
7	Q3V317 IDA1090.3961	Eph-rcpt lig (red) FN III (blue)
7	Q3V1W9 IDA1090.3961x2	Eph-rcpt lig (red) FN III (blue)
7	Q15375 IDA1090.3961x2.719[1245].11510[1660]	Eph-rcpt lig (red) FN III (blue) Prot kinase (orange) Tyr kinase (green) SAM 2 (yellow) SAM (teal)
6	Q16268 IDA1090.3961x2.1245[719]	Eph-rcpt lig (red) FN III (blue) Tyr kinase (green) Prot kinase (orange)
5	Q13146 IDA1090.3961x2.1245[719].1660[11510]	Eph-rcpt lig (red) FN III (blue) Tyr kinase (green) Prot kinase (orange) SAM (teal) SAM 2 (yellow)
3	Q5FW99 IDA1090.2049.3961x2.1245[719].1660	Eph-rcpt lig (red) EGF laminin (cyan) FN III (blue) Tyr kinase (green) Prot kinase (orange) SAM (teal)
3	Q4XMK2 IDA1090.3961x2.719[1245]	Eph-rcpt lig (red) FN III (blue) Prot kinase (orange) Tyr kinase (green)
1	Q6P9I9 IDA1090.2049.3961x2	Eph-rcpt lig (red) EGF laminin (cyan) FN III (blue)
1	Q6PVRS IDA1090.2049.3961x2.719[1245].1660	Eph-rcpt lig (red) EGF laminin (cyan) FN III (blue) Prot kinase (orange) Tyr kinase (green) SAM (teal)
1	Q6PFV6 IDA1090.3961.719[1245].1660	Eph-rcpt lig (red) FN III (blue) Prot kinase (orange) Tyr kinase (green) SAM (teal)
1	P29321 IDA1245[719].1660	Tyr kinase (green) Prot kinase (orange) SAM (teal)

View architectures of proteins containing a signature

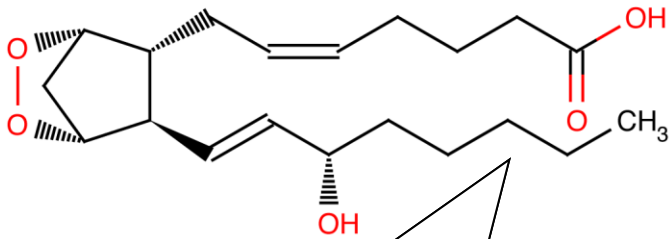
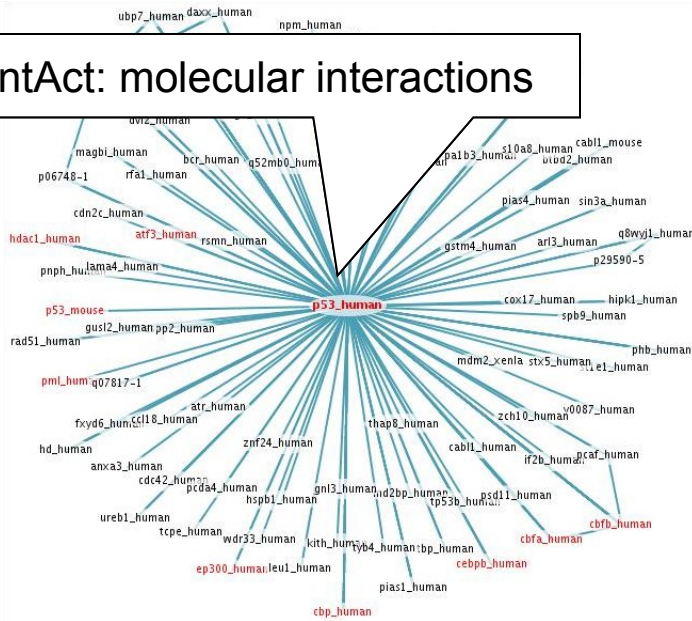


Visualize the taxonomic range for a protein signature



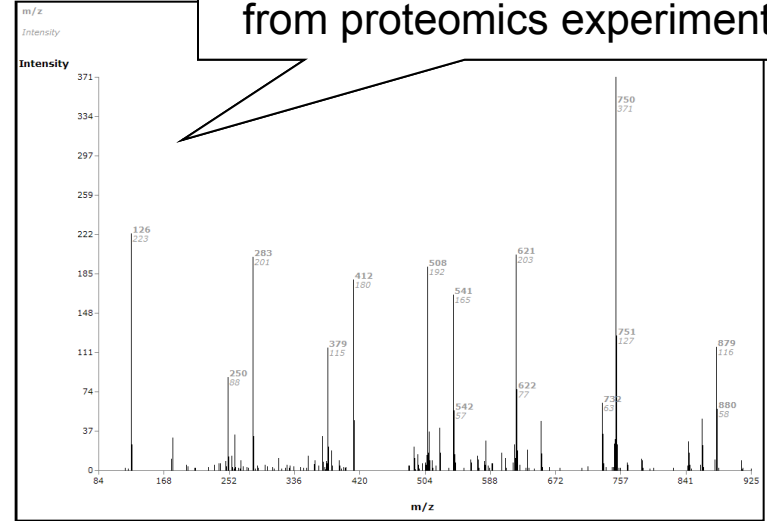
Proteomics services

IntAct: molecular interactions



ChEBI: small molecules

PRIDE: protein identifications from proteomics experiments



- EC 5 Isomerases
- EC 5.3 Intramolecular Oxidoreductases
- EC 5.3.99 Other Intramolecular Oxidoreductases
- EC 5.3.99.4 Prostaglandin-H synthase**

IntEnz view NC-IUBMB view ENZYME view

IntEnz Enzyme Nomenclature

EC 5.3.99.4

Names

Accepted name: prostaglandin-H synthase

Other name(s): PGI₂ synthase
PGI₂ synthetase
prostacyclin synthase
prostacycline synthetase
prostaglandin I₂ synthetase

Systematic name: (5Z,13E)-(15S)-9α,11α-epidioxy-15-hydroxyprosta-5,13-dienoic acid 6-isomerase

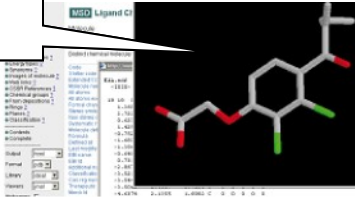
Reaction

(5Z,13E)-(15S)-9α,11α-epidioxy-15-hydroxyprosta-5,13-dienoate + H₂O → 9α-epoxy-11α,15-dihydroxyprosta-5,13-dienoate + H₂

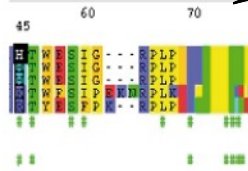
INTENZ: enzyme classification

Structures: PDBe

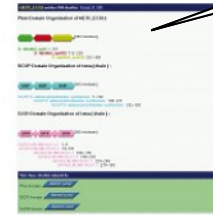
Ligands



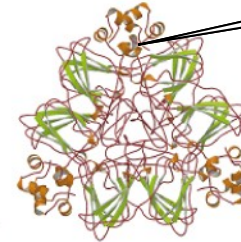
Sequence mapping



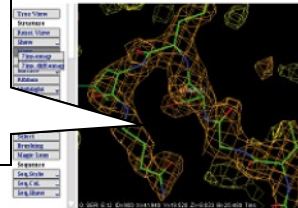
Linking to domain data



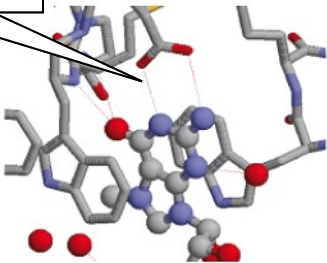
Assemblies



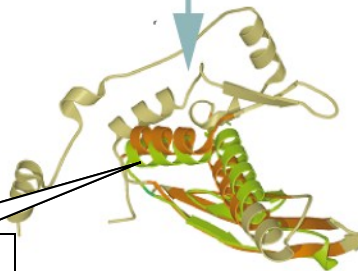
Electron density visualization



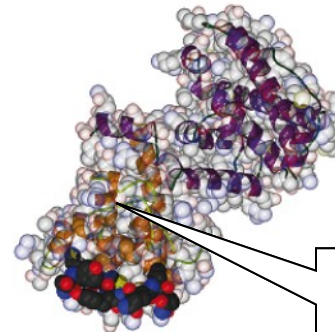
Active sites



Fold matching



Surface matching



Pathways: Reactome

Select a pathway

Reactome - a curated knowledgebase of biological pathways

The data displayed is for: **Homo sapiens** Use the menu to change the species. Check for cross-species comparison.

Reaction → Experimentally confirmed reaction → Manually inferred reaction → Electronically inferred reaction → Linked reactions →

Apoptosis	Biological oxidations	Botulinum neurotoxicity	Cell Cycle
Cell Cycle, Mitotic	DNA Repair	DNA Replication	Diabetes
Electron Transport Chain	Gap junction trafficking and regulation	Gene Expression	HIV
Hemostasis	Influenza Infection	Integration of energy metabolism	Integrin cell s
Lipid and lipoprotein metabolism	Membrane Trafficking	Metabolism of amino acids	Metabolism
Metabolism of nitric oxide	Metabolism of non-coding RNA	Metabolism of polyamines	Metabolism of vi
mRNA Processing	Nucleotide metabolism	Porphyryn metabolism	Pyruvate metat
Post-translational protein modification	Regulation of beta-cell development	Regulatory RNA pathways	Signal
Signaling by EGFR	Signaling by FGFR	Signaling by GPCR	Signaling in
Signaling by Insulin receptor	Signaling by NGF	Signaling by Notch	Opioic
Signaling by Rho GTPases	Signaling by TGF beta	Signaling by VEGF	Signa
Synaptic Transmission	Telomere Maintenance	Transcription	Tra
Transmembrane transport of small molecules			

Compare events in different species

Organism: Homo sapiens

Cellular compartment: cell

Represents GO biological process: insulin receptor signaling pathway [GO](#)

Equivalent event(s) in other organism(s):

- Insulin receptor mediated signaling [Drosophila melanogaster]
- Insulin receptor mediated signaling [Mus musculus]
- Insulin receptor mediated signaling [Rattus norvegicus]
- Insulin receptor mediated signaling [Gallus gallus]
- Insulin receptor mediated signaling [Tetraodon nigroviridis]
- Insulin receptor mediated signaling [Arabidopsis thaliana]
- Insulin receptor mediated signaling [Oryza sativa]
- Insulin receptor mediated signaling [Entamoeba histolytica]
- Insulin receptor mediated signaling [Dictyostelium discoideum]
- Insulin receptor mediated signaling [Plasmodium falciparum]

Link to source databases

Participating molecules

- 143B protein [cytosol] [UEGKRRHHURR](#)
- 3',5'-Cyclic AMP [cytosol] [CCP](#)
- 3-phosphoinositide dependent protein kinase-1 [plasma membrane] [UEGKU](#)
- 40S small ribosomal protein 6 [cytosol] [UEGRRRRRHHHHHUR](#)
- 4E-BP [cytosol] [UEGKRHHUR](#)
- 4E-BP1-P [cytosol] [UEGKRHHUR](#)
- Activated PI3K [endosome membrane, plasma membrane]
- ADP [cytosol] [CCP](#)
- AMP [cytosol] [CCP](#)
- ATP [cytosol] [CCP](#)
- ...

List all 96 participating molecules

Export pathway to your favourite modelling software

Diagram

```

    graph LR
      A[Re-integration of insulin receptor into plasma membrane] --> B[Insulin receptor [plasma membrane]]
      B --> C[Insulin [extracellular region]]
      C --> D[Insulin binding [Homo sapiens]]
      D --> E[Insulin:insulin receptor [plasma membrane]]
      E --> F[autophosphorylation of insulin receptor]
  
```

Details

Stable Identifier: REACT_1459.2

Author: .2003-07-28

Under normal physiological conditions blood glucose levels are kept under tight control by a series of regulated steps that ensure glucose homeostasis. Upon feeding glucose levels rise and in response to this the body secretes insulin from pancreatic beta-cells into the blood. At physiological concentrations insulin is present in the blood in its monomeric form. Binding of insulin to its receptor occurs on the receptor alpha-subunits. There are two binding domains involved on the receptor (L1 and L2) and it is thought that the amino-terminus of insulin binds with L1 on one of the alpha-subunits and the carboxyterminus with L2 on the other alpha-subunit.

The binding of insulin to its receptor causes a conformational change in the alpha-subunits. This in turn produces a conformational change in the beta-subunits leading to the activation of the intrinsic insulin receptor tyrosine kinase.

View reactions and events in detail

User support

- 2Can bioinformatics user support – www.ebi.ac.uk/2Can
- Online help pages – www.ebi.ac.uk/help
- E-mail support – www.ebi.ac.uk/support



Research

www.ebi.ac.uk/groups

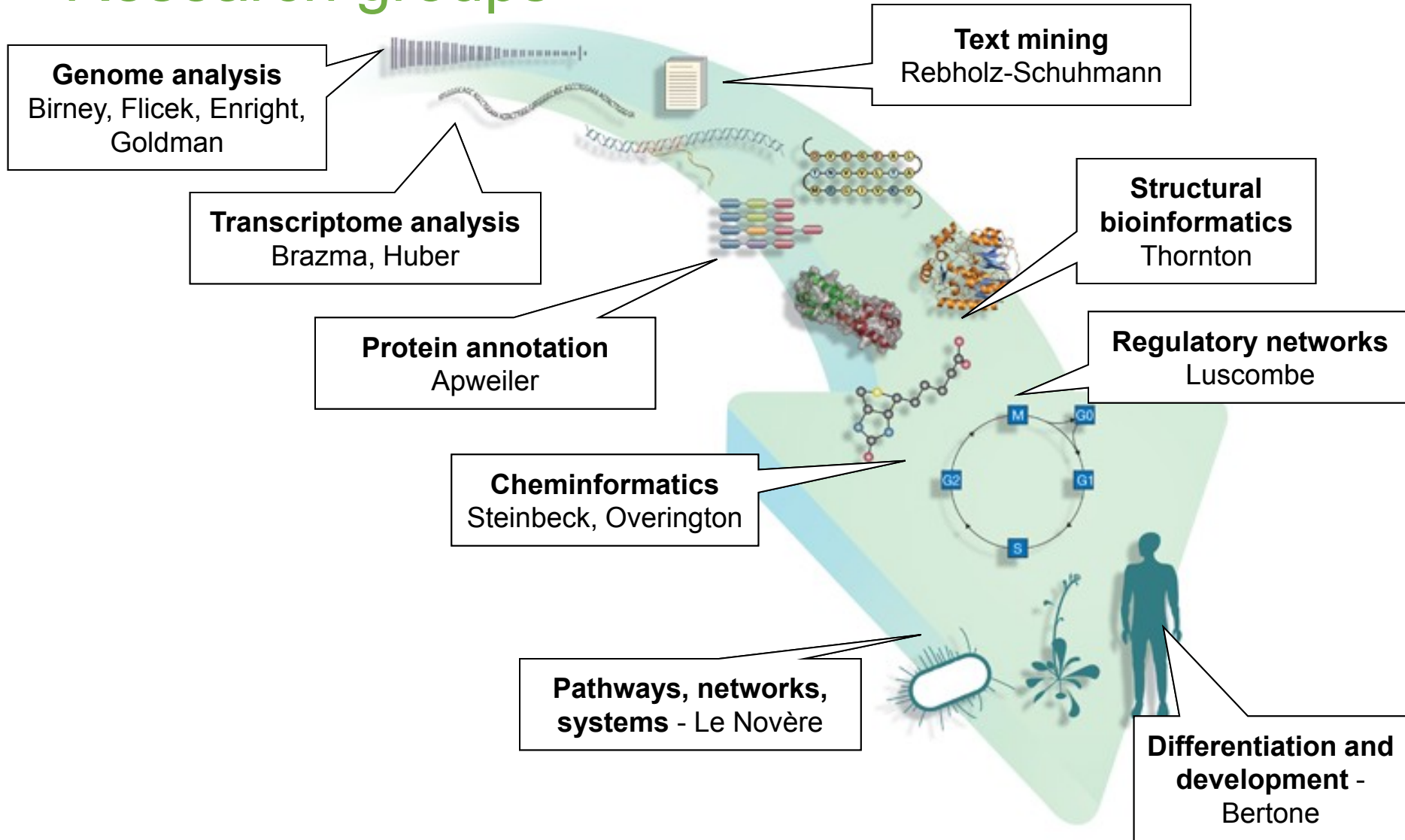
EMBL-EBI

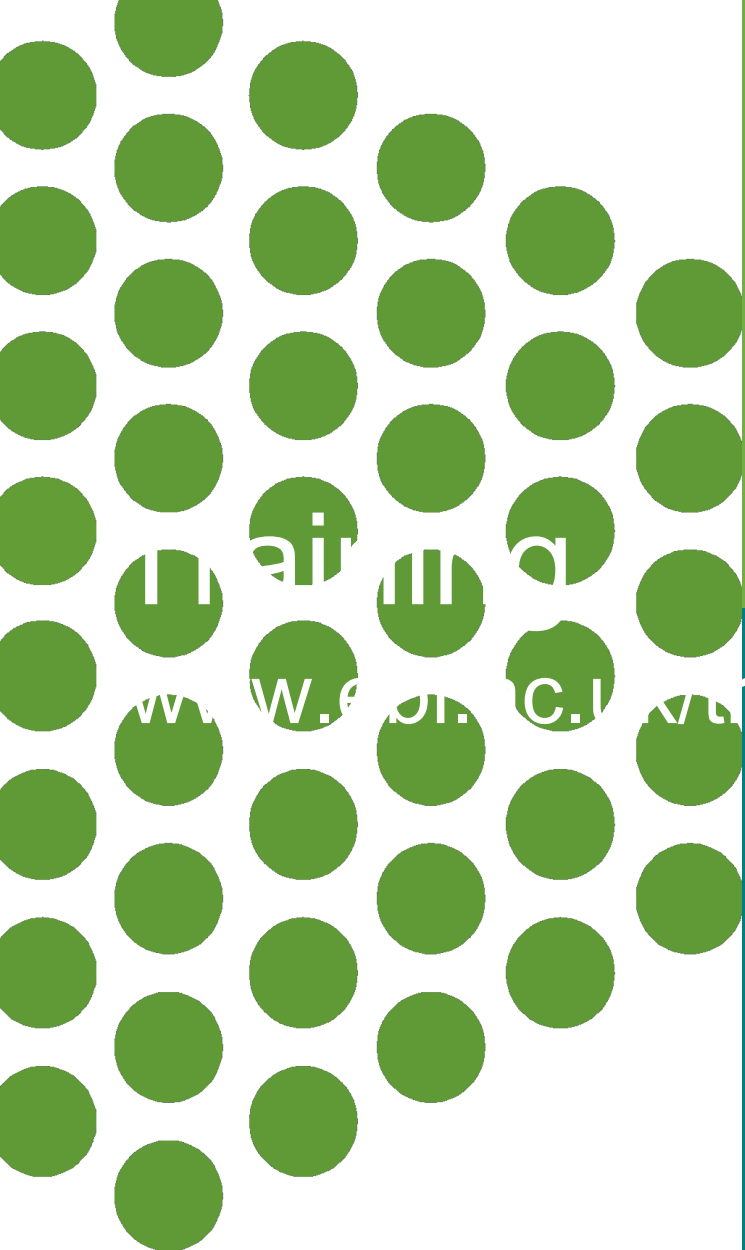


Key facts about research

- The EBI provides a unique environment for bioinformatics research
- Eight dedicated research groups aim to understand biology through new approaches to interpreting biological data
- Services teams also carry out R&D to enhance existing services and develop new ones
- Research programme complements services and the two are mutually supportive

Research groups



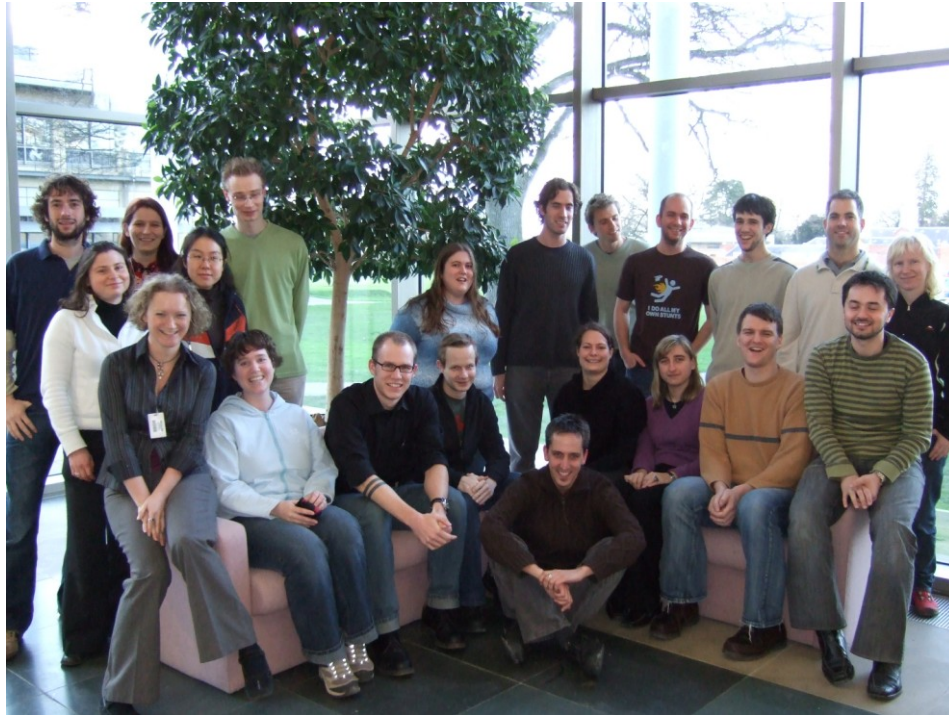


Training
www.ebi.ac.uk/training

EMBL-EBI



Predoc and postdoc training

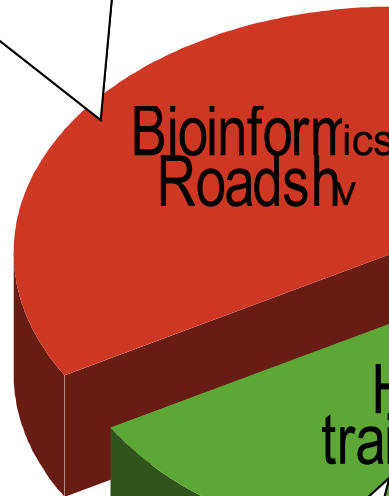


- Annual Open Days for bioinformatics masters' students
- PhD studentships through EMBL International PhD Programme
- Short-term placements for visiting PhD students through EU-funded Marie Curie Fellowships

A tripartite user-training programme

Training comes to you

www.ebi.ac.uk/training/roadshow



Bioinformatics
Roadshow

Training any time, anywhere, at any pace

www.ebi.ac.uk/training/elearning

eLearning
programme

Hands-on
training at
EMBL
EBI

Hands-on user training on all our core data resources for lab-based researchers

www.ebi.ac.uk/training/handson

Hands-on training for all levels of experience

- **Interactive training** in our purpose-built IT training suite at EMBL-EBI, Hinxton, Cambridge
- **Learn from the EBI's experts** through a combination of talks and practical exercises
- Take a **tour** of all our core data resources, or **focus in on specific data types**
- Full programme at www.ebi.ac.uk/training/hands-on



<http://www.ebi.ac.uk/training/handson/>

Hands-on training at EBI



Interactions, Pathways & Networks

Title: Interactions, Pathways & Networks

Date: 15th-18th Jun 2009

Venue: Genome Campus, Hinxton, Cambridge, UK

Registration Deadline: 18th May 2009

[more info...](#)



Genomics, proteomics, transcriptomics, protein structures...

What our trainees say...

“ this course gave me just what I was looking for
From sequence to gene ”

“ superb course ran by attentive tutors
” Proteomes

“ very nice to hear about tools that biologists usually are not aware of
Protein to Proteomes ”

“ A very valuable experience. I'll definitely tell my colleagues about EBI's courses
Protein to Proteomes ”

“ it's been a great learning experience
” From sequence to gene

“ Great facilities, very good presentations, interesting content
” Transcriptomics

“ the hands on sessions were clear
” Transcriptomics

“ the best I have attended
” Proteomes

Moodle-based eLearning platform



www.ebi.ac.uk/training/elearning

Welcome to the EBI Staging Moodle Installation

My courses

EBI and EB-eye
Administrator: [The Consultants-E](#)
Administrator: [Nicky Hockly](#)
Course creator: [EBI Student](#)
Administrator: [EBI Trainer](#)
This course provides an introduction to the EBI and shows you how to use the EB-eye global search engine. It is also the place to go for help and advice with this and the other EBI Training courses. You will also find forums where you can interact with other EBI Training users.

Sequence Searching
Administrator: [The Consultants-E](#)
Administrator: [Nicky Hockly](#)
Course creator: [EBI Student](#)
Administrator: [EBI Trainer](#)
This course will teach you how to perform sequence similarity searches on the EBI databases. Here, you can learn how to use the basic local alignment search tool BLAST, learn how to change BLAST's parameters to optimise your search, learn about other alignment tools (such as FASTA, MPSrch and SSAHA) and when you might want to use them, find out how to search for patterns, profiles and Hidden Markov Models in a sequence, and learn to build your own profiles using PSI-BLAST. You also have access to a Glossary of key terms associated with sequence searching.

Patent Searching
Administrator: [The Consultants-E](#)
Administrator: [Nicky Hockly](#)
Course creator: [EBI Student](#)
Administrator: [EBI Trainer](#)
This course covers patent searching in the EBI databases. You will learn about the patent resources available at the EBI, how to search for full-text patents, how to use text- and sequence-based search methods to find patents relevant to a particular gene or protein, how to use the EBI's sequence version archives to find the original sequences associated with a patent, and how to download patent sequences.

Literature Searching
Administrator: [The Consultants-E](#)
Administrator: [Nicky Hockly](#)
Course creator: [EBI Student](#)
Administrator: [EBI Trainer](#)
This course will help you to make the most of the EBI's literature resources. You will learn how to use CiteXplore, the EBI's portal to the literature, and find out which databases you can search using CiteXplore. You will also learn how to use text-mining (or 'text mining') tools Whatizit, EBIMed and Protein Corral. You also have access to a Glossary of key terms associated with literature searching.

Turn editing on

Activities

Forums

Calendar

May 2007

Sun	Mon	Tue	Wed	Thu	Fri	Sat
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

Courses available

- EBI and EB-eye
- Sequence searching
- Patent searching
- Literature searching
- Ensembl
- Transcriptomics

Each course is modular

A course contains 3–5 modules (~30 min each)

Each module contains...

The screenshot shows a course module titled "2 BLAST for Beginners". The text describes the section's focus on using the BLAST search tool. Below the text, it lists learning objectives and provides links to a video tutorial, a print tutorial, a key concepts quiz, and a reflective task. Four callout boxes on the right side of the page point to these specific elements, describing the learning method for each: video tutorial (watching and listening), print tutorial (reading), quiz (testing understanding), and reflective task (practicing).

2 BLAST for Beginners

This section looks at using the general search tool BLAST to do a basic nucleotide or protein-based similarity search. **BLAST** stands for **B**asic **L**ocal **A**lignment **S**earch **T**ool. We recommend that you work through this section of the course if you do not have much experience of BLAST.

You will learn about:

- the basics of BLAST – what it is and when to use it.
- how BLAST works.
- how to run a simple BLAST search.
- how to interpret the results of a search.
- how to interpret an alignment in BLAST.

[BLAST for Beginners - Video Tutorial](#)

[BLAST for Beginners - Print Tutorial](#)

[BLAST for Beginners - Key Concepts Quiz](#)

[Reflective Task](#)

Video tutorial
learn by watching and listening

Print tutorial
Learn by reading

Quiz
Learn by testing your understanding

Reflective task
Learn by practicing

Please beta-test and provide feedback!

industry support
www.ebi.ac.uk/industry

EMBL-EBI

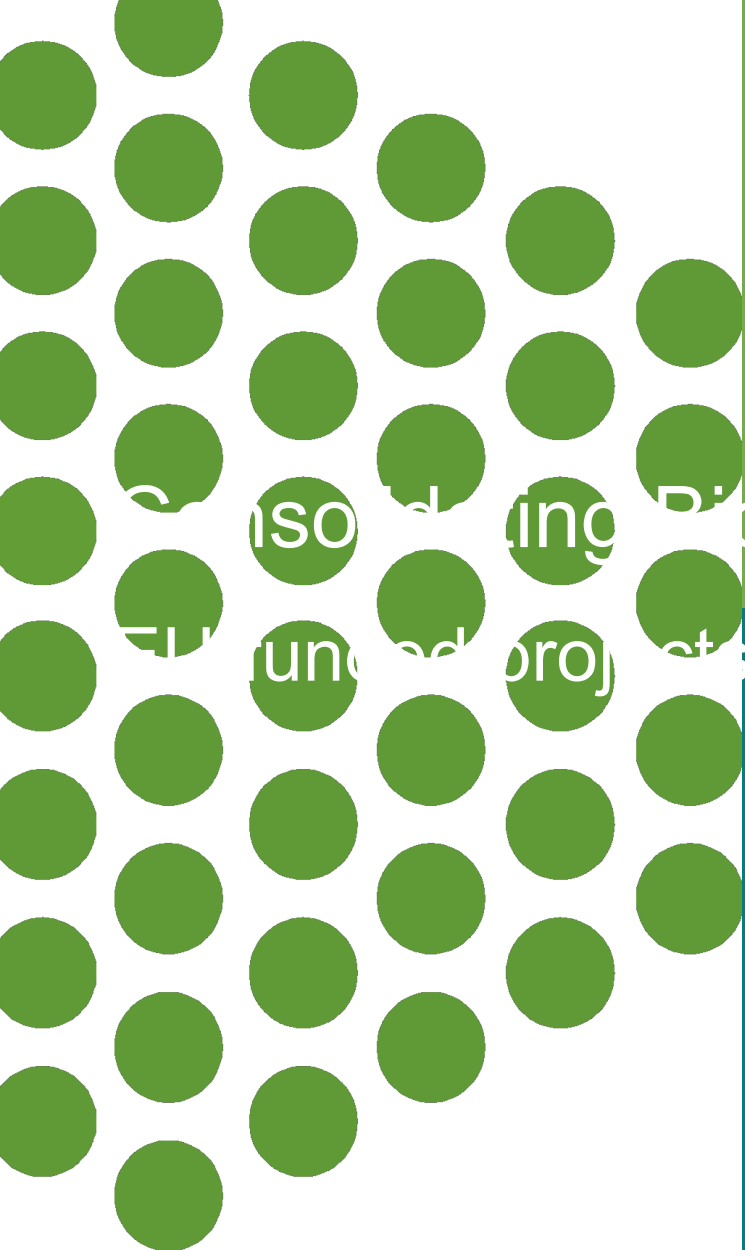


The EBI Industry Programme

- Enables industry to adapt quickly to, and maximise the benefit from, innovations in bioinformatics.
- Membership benefits include:
 - Research of benefit to industry
 - Expert training
 - Standards development
 - Technical development
 - Networking opportunities
- Membership is by invitation and members subscribe on an annual basis

Industry Programme members

- AstraZeneca
- Bayer Schering Pharma AG
- Boehringer Ingelheim Pharma GmbH & Co. KG
- Eli Lilly & Company
- Galderma
- GlaxoSmithKline
- F. Hoffman-La Roche
- Johnson & Johnson Pharmaceutical Research & Development
- Merck KGaA
- Nestlé Research Centre
- Orion Pharma
- Philips Research
- Pfizer Ltd
- Syngenta Limited
- Sanofi-Aventis Recherche & Développement
- Unilever



Coordinating Bioinformatics in Europe

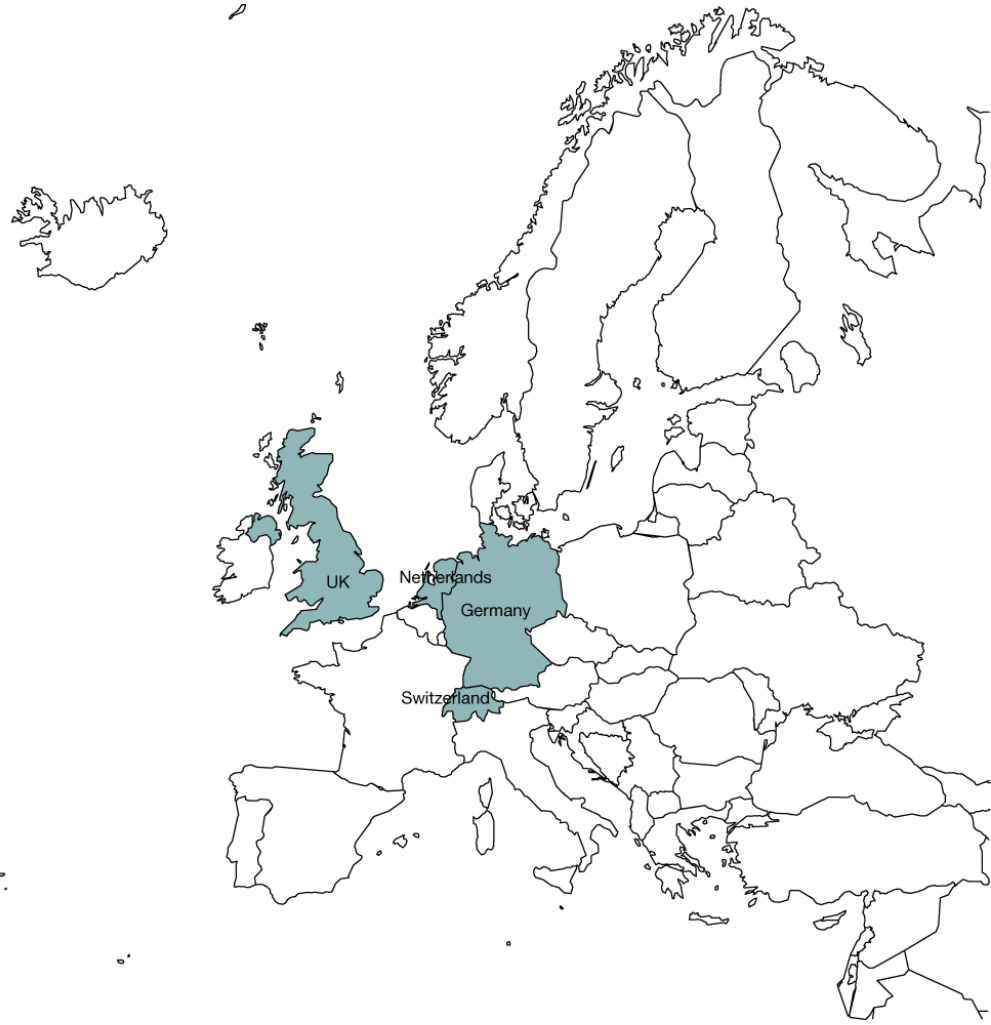
EU funded projects coordinated by the EBI

EMBL-EBI



SLING – Serving life science information in the next generation

- Providing unrestricted access to some of the world's most important biological databases
- Bioinformatics roadshows provide hands-on training for users
- Funded by the European Commission within its FP7 Programme within the Research Infrastructure Programme
- 4 partners in 4 countries



ELIXIR – European life sciences infrastructure for biological information

To build a sustainable European infrastructure for biological information supporting life science research and its translation to:

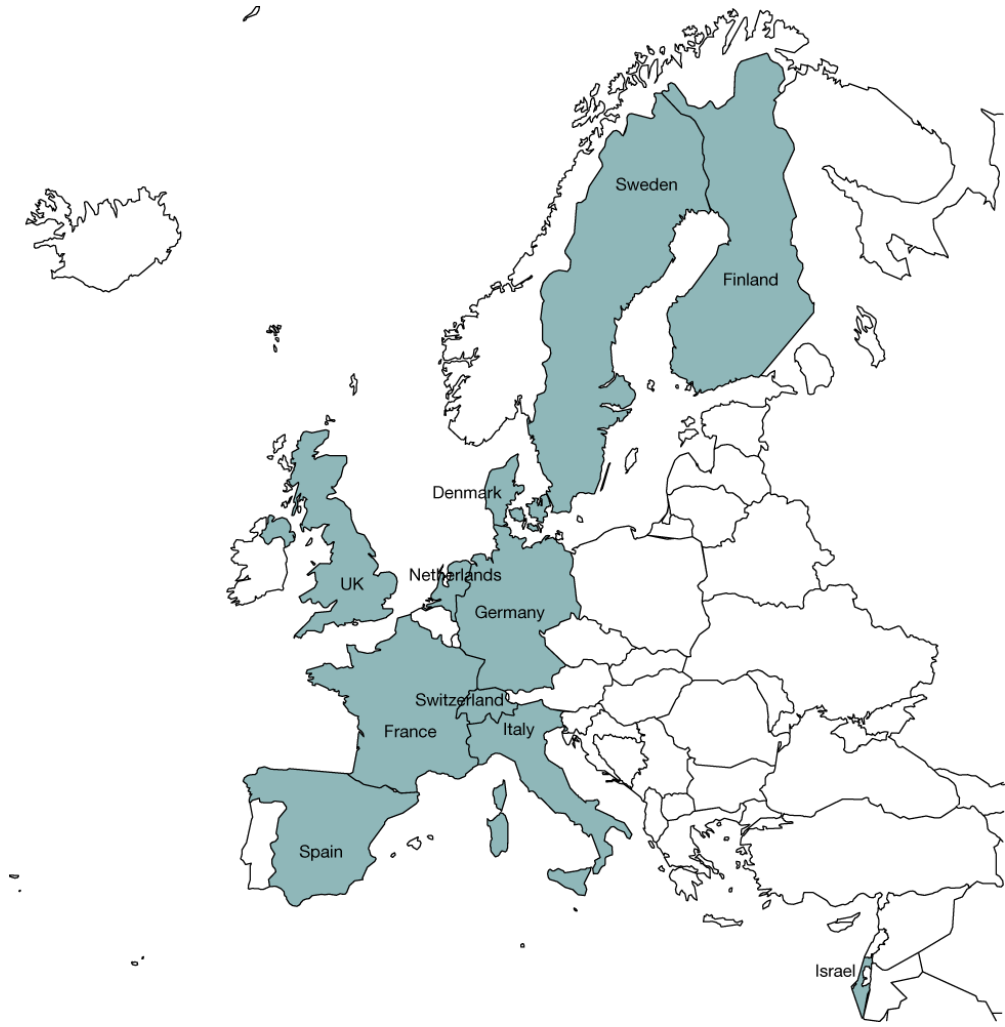
- medicine,
- the environment,
- the bioindustries, and
- society

32 participants in 13 countries



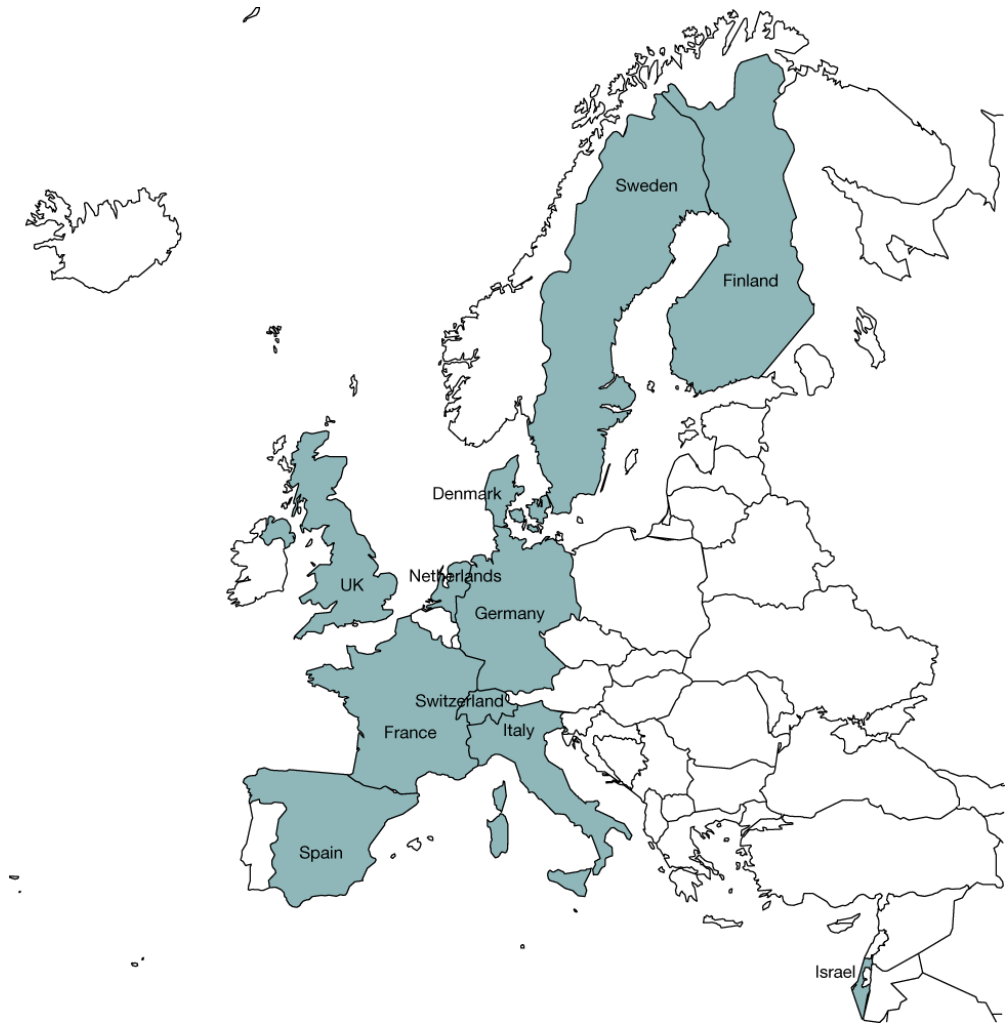
ENFIN Network of Excellence

- Brings together experimentalists and computational biologists to develop the next generation of informatics resources for systems biology
- Funded by the European Commission within its FP6 programme under the thematic area 'Life sciences, genomics and biotechnology for health'
- 20 partners in 13 countries
- **www.enfin.org**



EMBRACE Network of Excellence

- Aims to enable bioinformatics research through better interoperability of servers, databases and services
- Funded by the European Commission within its FP6 programme under the thematic area 'Life sciences, genomics and biotechnology for health'
- 17 partners in 11 countries
- **www.embracegrid.info**



BioSapiens Network of Excellence

- A large-scale, concerted effort to annotate genome data through a virtual institute for genome annotation and a European School of Bioinformatics
- Funded by the European Commission within its FP6 programme under the thematic area 'Life sciences, genomics and biotechnology for health'
- 24 partners in 14 countries
- **www.biosapiens.info**

