

EMBL-European Bioinformatics Institute

What is the European Bioinformatics Institute?

- Part of the European Molecular Biology Laboratory
- International, non-profit research institute
- Europe's hub for biological data services and research



The five branches of EMBL

Heidelberg



- Basic research in molecular biology
- Administration
- EMBO

Hamburg



Structural biology

Hinxton



Bioinformatics

Grenoble



Structural biology

Monterotondo



Mouse biology

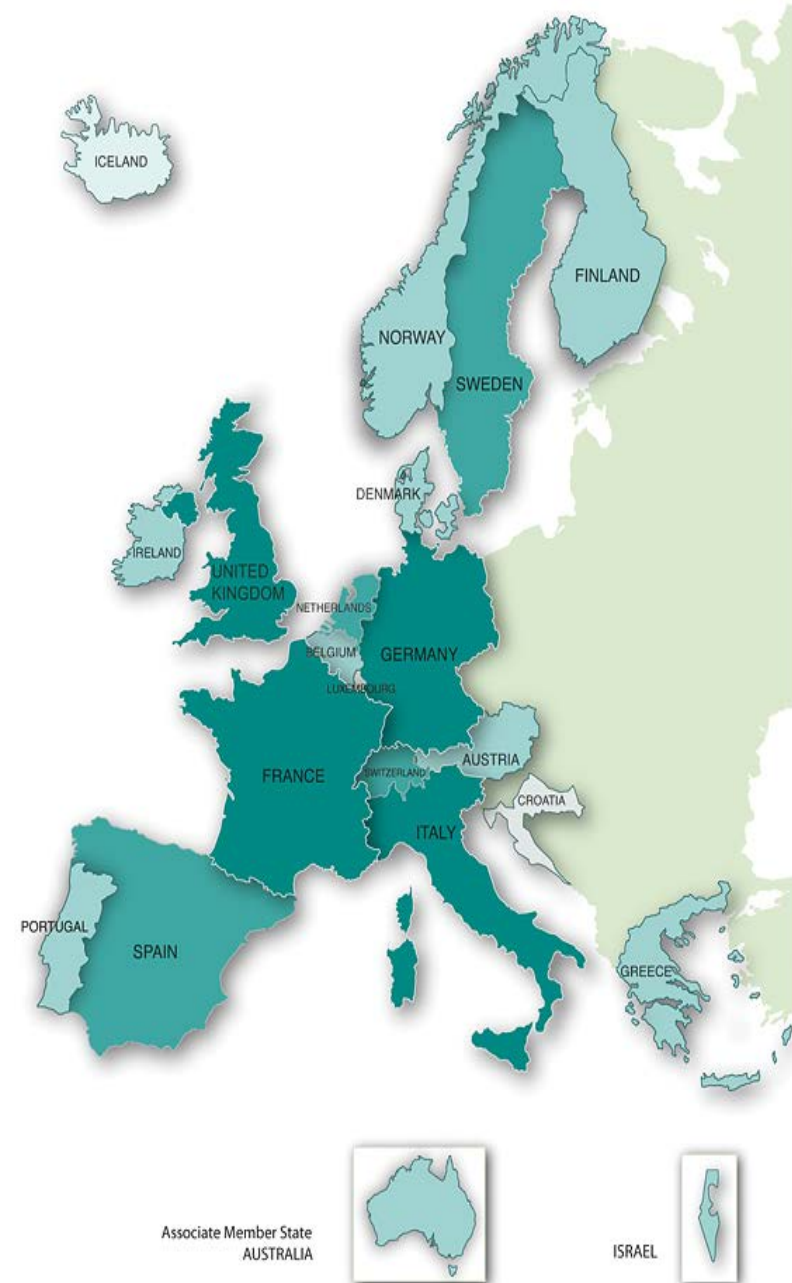
• 1,500 staff

• >60 nationalities

EMBL member states

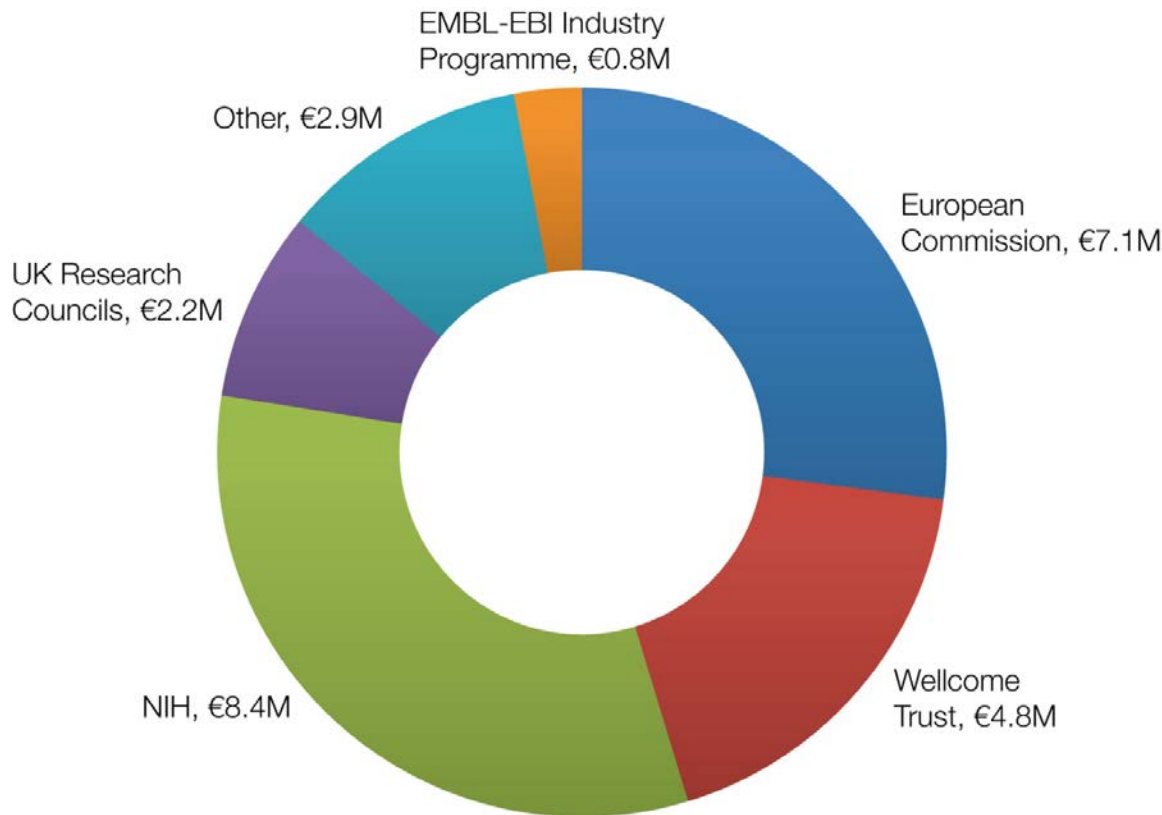
Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom

Associate member state: Australia



EMBL-EBI also has significant external funding

Sources of external funding for the year 2011. The Wellcome Trust also supports us through provision of our buildings.



The UK government granted £75 million in 2011 to support our role in coordinating ELIXIR, the emerging infrastructure for biological data in Europe.

The Wellcome Trust Genome Campus



Data centre

Sanger labs
/ informatics






Cairns
Pavilion
(shared)

Sanger Institute
Sulston building

EMBL-EBI

© John Freebury

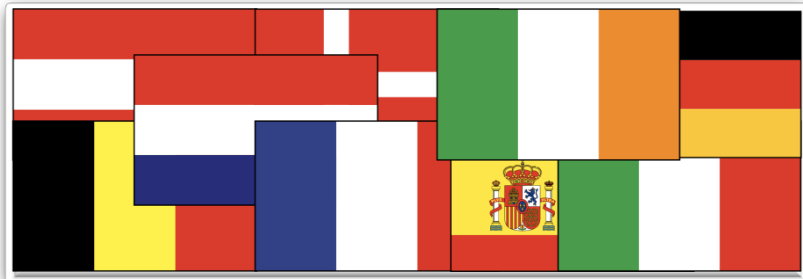
EMBL-EBI's mission

-  Provide **freely available data** and bioinformatics **services** to all facets of the scientific community in ways that promote scientific progress
-  Contribute to the advancement of biology through basic investigator-driven **research** in bioinformatics
-  Provide advanced bioinformatics **training** to scientists at all levels, from PhD students to independent investigators
-  Disseminate cutting-edge technologies to **industry**
-  **Coordinate** biological data provision across Europe

The people behind the data and research



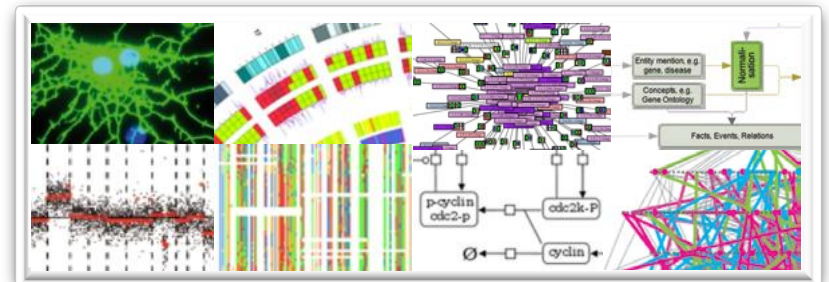
~550 staff



>45 nationalities



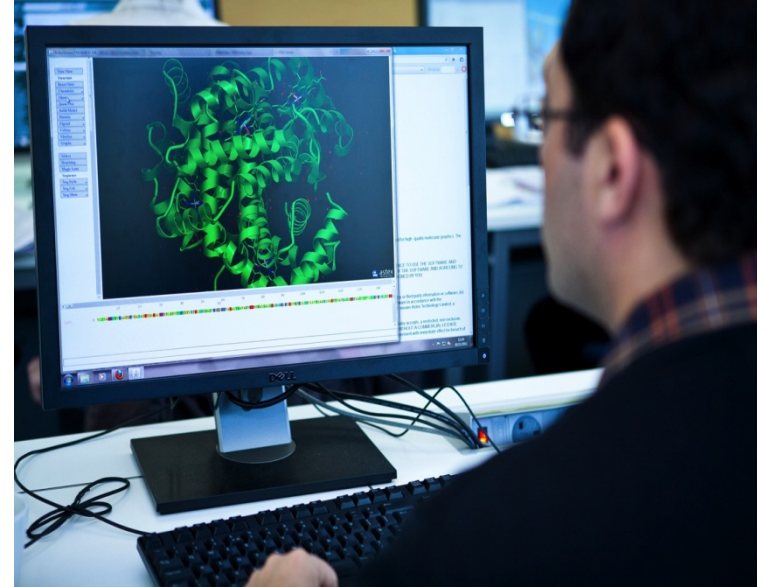
~400 in services & support



~100 in research

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

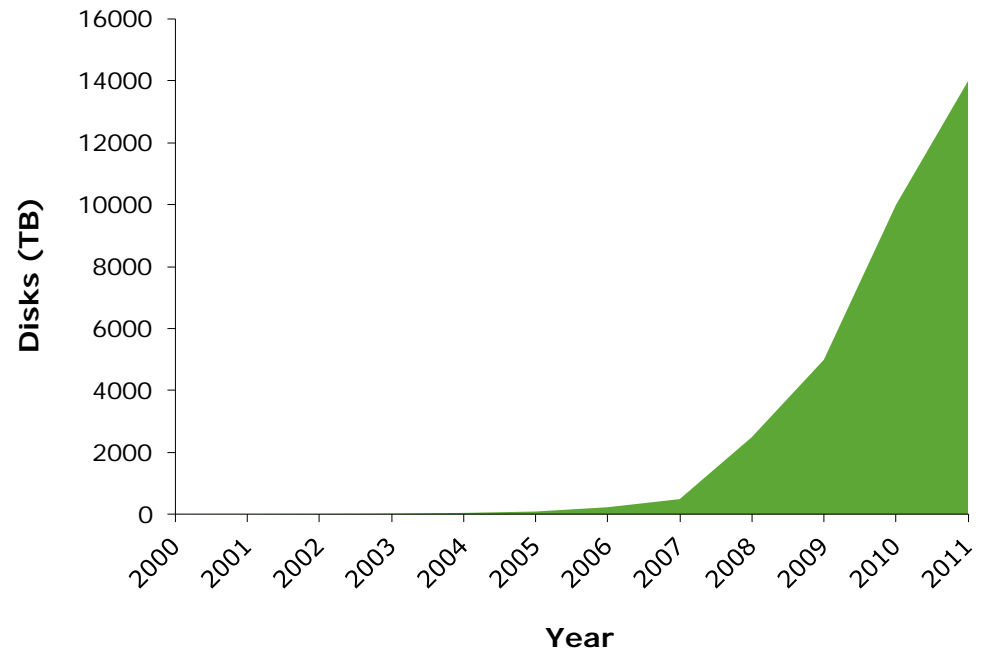


© Samuel Kerrien

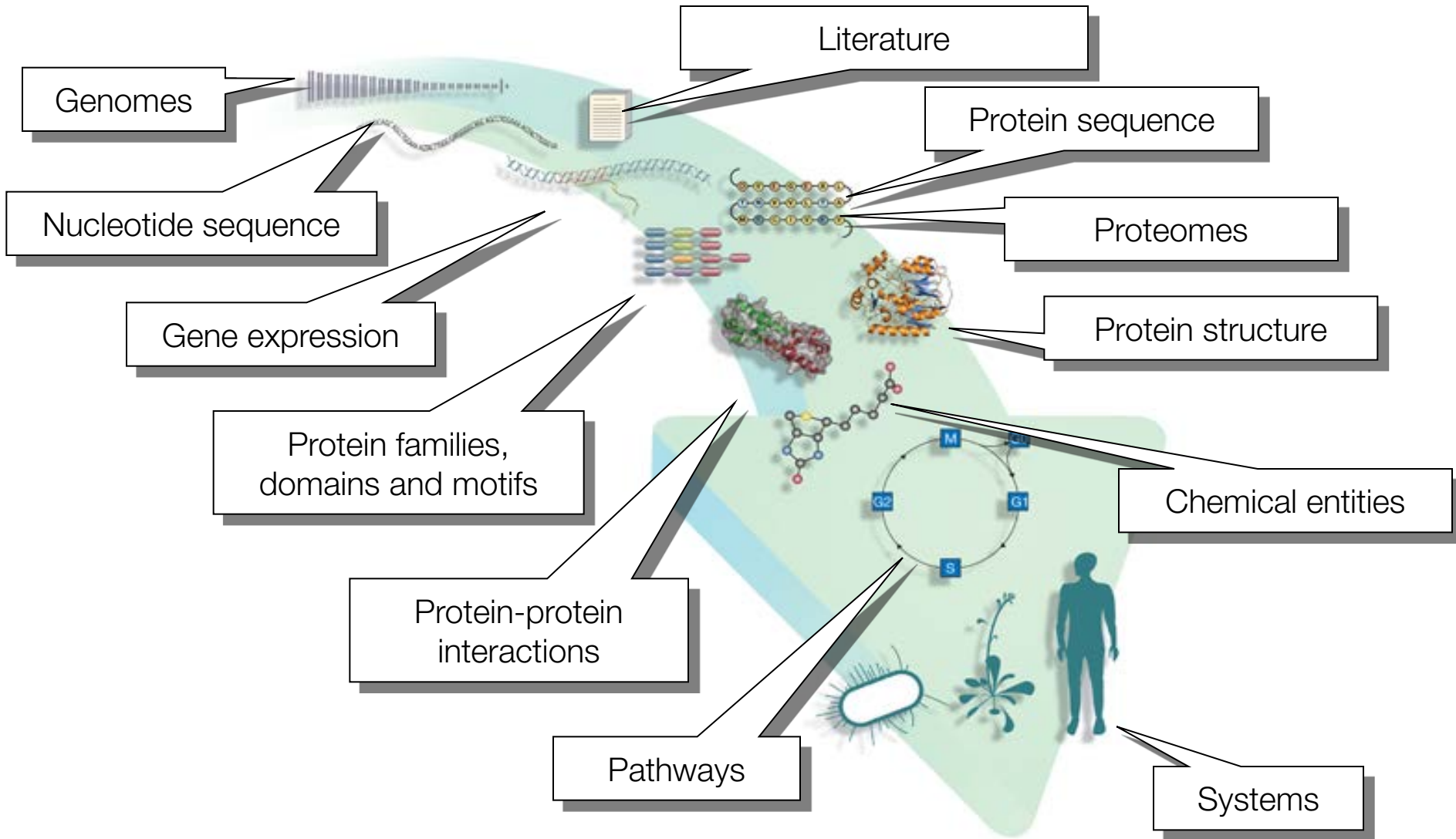
Biology is changing

- Data explosion
- New types of data
- High-throughput biology
- Emphasis on **systems**, not reductionism
- Growth of applied biology
 - molecular medicine
 - agriculture
 - food
 - environmental sciences

Growth of raw storage at EMBL-EBI (in terabytes)



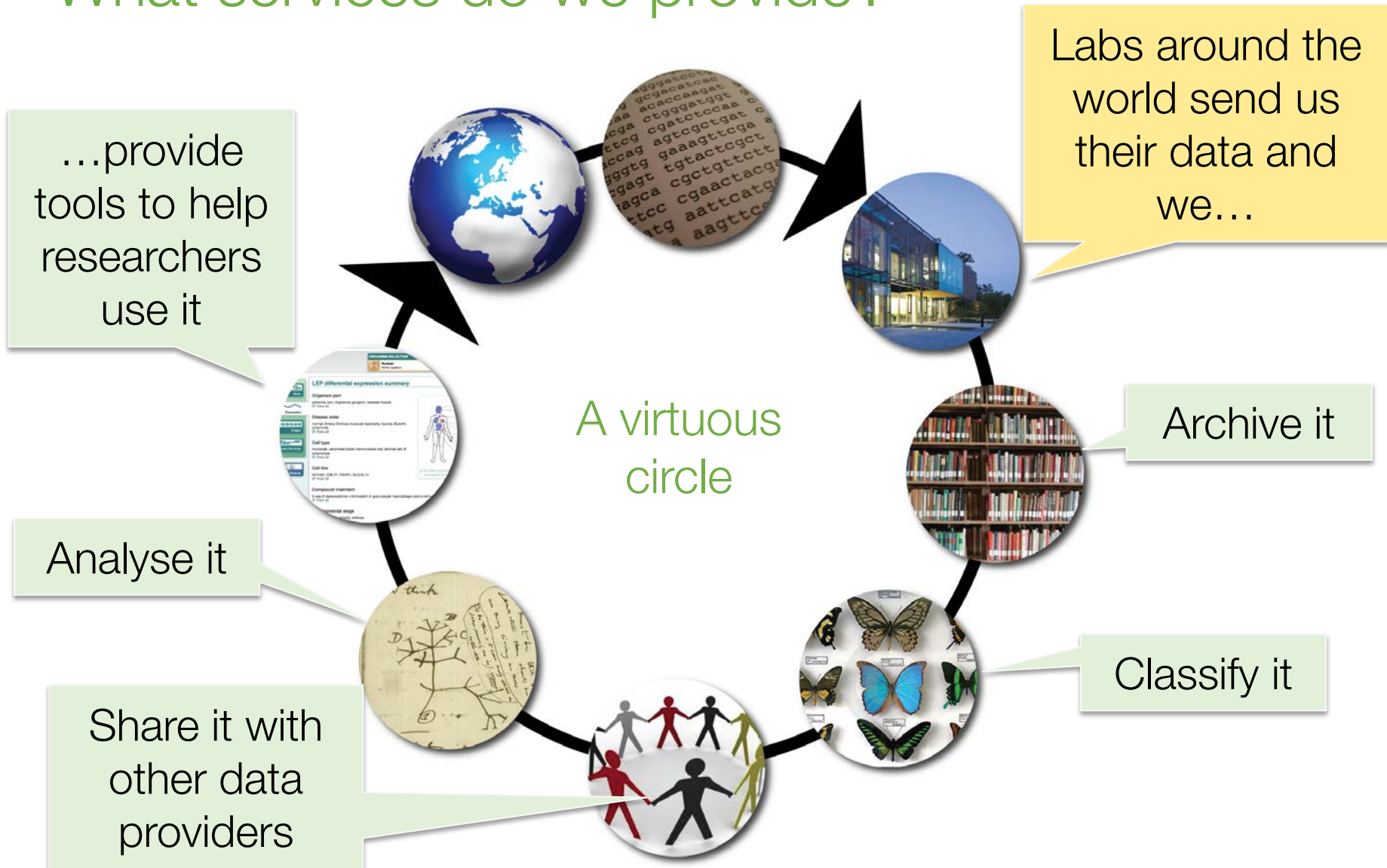
Different Types of Biological Data



Bioinformatics services

www.ebi.ac.uk/services

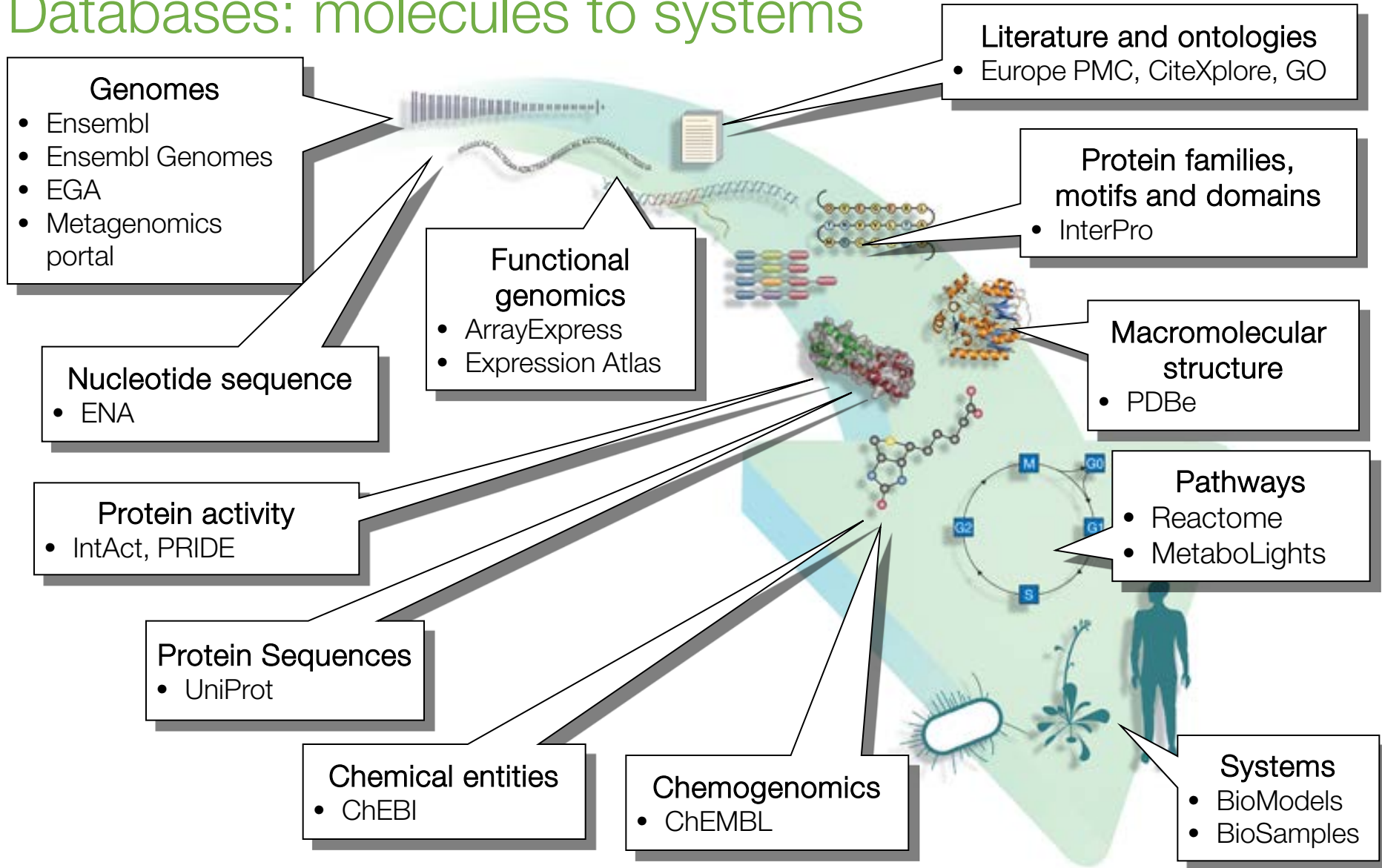
What services do we provide?



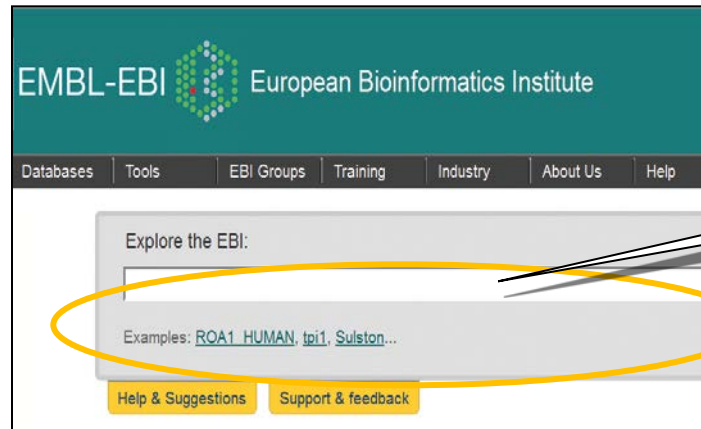
Key facts about EBI services

- **Freely available**
- The world's most **comprehensive** collection of molecular databases
- Globally **coordinated** data collection and dissemination
- Produced in **collaboration** with other world leaders, e.g.
 - NCBI, USA
 - National Institute of Genetics, Japan
 - Swiss Institute of Bioinformatics (SIB)
 - Cold Spring Harbor Laboratory, USA

Databases: molecules to systems



Search the data: www.ebi.ac.uk



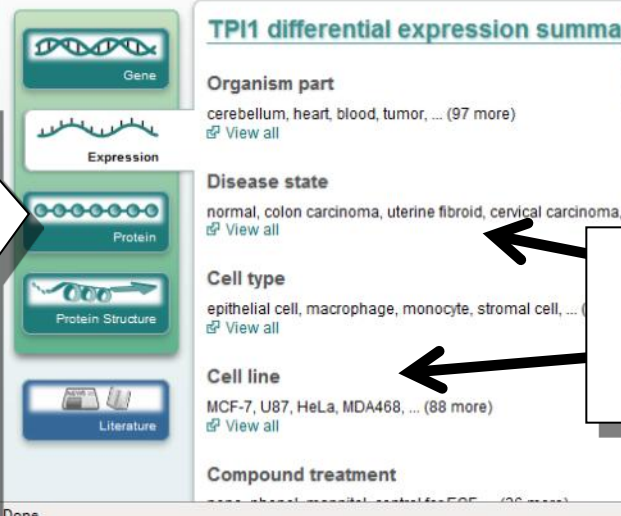
Access from the EBI's homepage

Species selector allows for easy comparison

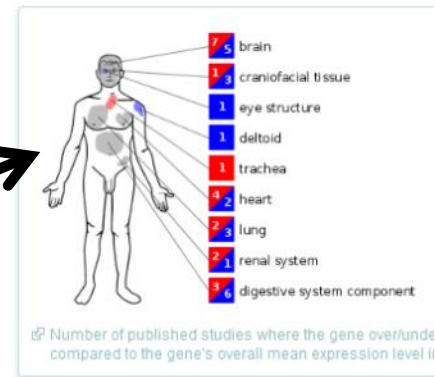


Data organised according to:

- gene
- expression
- protein
- structure
- literature



Explore data, return easily to your results



Genomes: *e!* Ensembl

- Explore vertebrate (and selected invertebrate) genomes
- Produced in collaboration with the WT Sanger Institute

Genomic alignments

Pick a genome

Chromosomes

Genes

Synteny

User Upload

Variations

Gene trees

Gene families

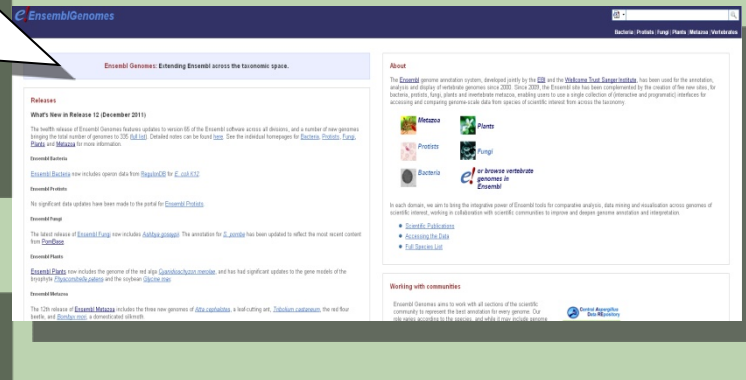
Variant Effect Predictor

1	881907	881906	-/C	+	
12-1017956	ENSG00000060237	ENST00000315938	NON_SYNONYMOUS_CODING	/C	+
12-1017956	ENSG00000060237	ENST00000252477	NON_SYNONYMOUS_CODING	A	+
12-1017956	ENSG00000060237	ENST00000349808	NON_SYNONYMOUS_CODING	/C	+
12-1017956	ENSG00000002016	ENST00000358495	DOWNSTREAM		-
12-1017956	ENSG00000002016	ENST00000228345	DOWNSTREAM		-
19	66520	66520	G/A		+
8	150029	150029	A/T		+

Genomes: *e!* Ensembl Genomes

- Explore genome-scale data from bacteria, protists, fungi, plants and invertebrate metazoa.

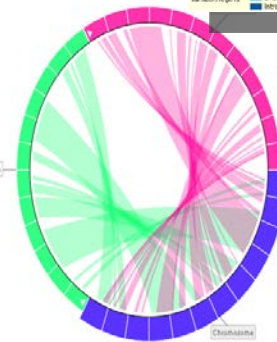
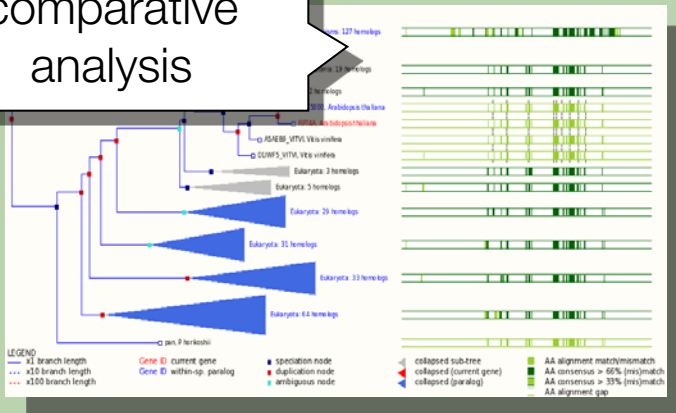
Genome portals for the five kingdoms of life



Variation data for plant, metazoan and fungal species



Pan-taxonomic comparative analysis

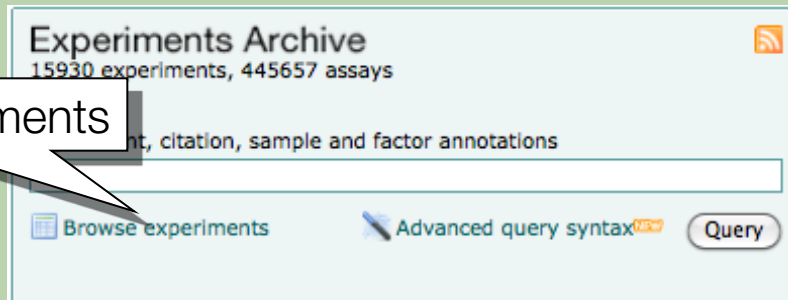


Multi-way comparison of whole bacterial chromosomes

Transcription: ARRAYEXPRESS

- Functional genomics experiments – RNA-Seq, ChIP-Seq and array-based technologies
- MIAME-standard compliant.

Browse experiments

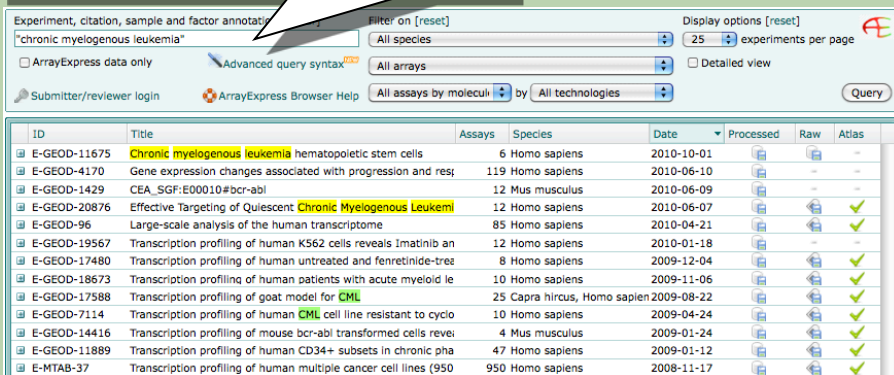


Experiments Archive
15930 experiments, 445657 assays

Experiment, citation, sample and factor annotations

[Browse experiments](#) [Advanced query syntax](#) [Query](#)

Search by keyword



Experiment, citation, sample and factor annotations

Filter on [reset] Display options [reset]

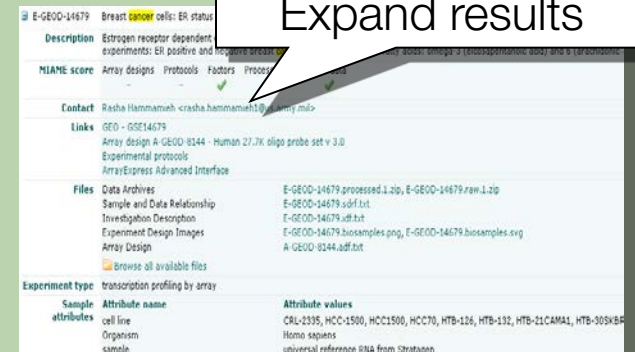
"chronic myelogenous leukemia" All species 25 experiments per page

ArrayExpress data only Advanced query syntax Detailed view

[Submitter/reviewer login](#) [ArrayExpress Browser Help](#) All assays by molecule by All technologies [Query](#)

ID	Title	Assays	Species	Date	Processed	Raw	Atlas
E-GEOD-11675	Chronic myelogenous leukemia hematopoietic stem cells	6	Homo sapiens	2010-10-01			
E-GEOD-4170	Gene expression changes associated with progression and res;	119	Homo sapiens	2010-06-10			
E-GEOD-1429	CEA_SGF:E00010#bcr-abl	12	Mus musculus	2010-06-09			
E-GEOD-20876	Effective Targeting of Quiescent Chronic Myelogenous Leukem	12	Homo sapiens	2010-06-07			
E-GEOD-96	Large-scale analysis of the human transcriptome	85	Homo sapiens	2010-04-21			
E-GEOD-19567	Transcription profiling of human K562 cells reveals Imatinib an	12	Homo sapiens	2010-01-18			
E-GEOD-17480	Transcription profiling of human untreated and fenretinide-tre	8	Homo sapiens	2009-12-04			
E-GEOD-18673	Transcription profiling of human patients with acute myeloid le	10	Homo sapiens	2009-11-06			
E-GEOD-17588	Transcription profiling of goat model for CML	25	Capra hircus, Homo sapien	2009-08-22			
E-GEOD-7114	Transcription profiling of human CML cell line resistant to cyclo	10	Homo sapiens	2009-04-24			
E-GEOD-14416	Transcription profiling of mouse bcr-abl transformed cells revei	4	Mus musculus	2009-01-24			
E-GEOD-11889	Transcription profiling of human CD34+ subsets in chronic pha	47	Homo sapiens	2009-01-12			
E-MTAB-37	Transcription profiling of human multiple cancer cell lines (950	950	Homo sapiens	2008-11-17			

Expand results



E-GEOD-14679 Breast cancer cells: ER status

Description Estrogen receptor dependent experiments: ER positive and negative breast cancer cell lines

MIAME score Array designs Protocols Factors Processed

Contact Rada Hanamamab <rada.hanamamab@ebi.ac.uk>


Links GEO - GSE14679
Array design A-GEOD-8144 - Human 27,7K oligo probe set v 3.0
Experimental protocols
ArrayExpress Advanced Interface

Files Data Archives E-GEOD-14679 processed.1.zip, E-GEOD-14679.raw.1.zip
E-GEOD-14679.pdf.txt
Investigation Description E-GEOD-14679.txt
Experiment Design Images E-GEOD-14679 biosamples.png, E-GEOD-14679 biosamples.svg
Array Design A-GEOD-8144.pdf.txt
Browse all available files

Experiment type transcription profiling by array

Sample attributes
Attribute name cell line Organism sample
Attribute values CRL-2335, HCC-1500, HCC1500, HCC70, HTB-116, HTB-132, HTB-21CAMAL1, HTB-3054EP, Homo sapiens
universal reference RNA from Stratagene

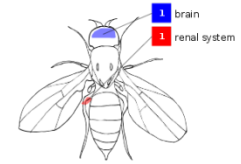
Read descriptions of sample properties



Experiment E-MEXP-2612
Transcription profiling by array of human Graves disease short and long course samples (8 samples)

Source Name	DiseaseState	Organism	Sex	Sample Characteristics			Factor Values				
				OrganismPart	DevelopmentalStage	BioSourceType	ClinicalHistory	DISEASE_STATE	CLINICAL_HISTORY	SEX	Array Data File
C1 TB209	Control	Homo sapiens	male	thyroid gland	adult	frozen_sample		Control		male	TB209.CEL
C2 TB377	Control	Homo sapiens	female	thyroid gland	adult	frozen_sample		Control		female	TB377.CEL
LC TB413	Graves' disease	Homo sapiens	female	thyroid gland	adult	frozen_sample	long course	Graves' disease	long course	female	TB413.CEL
LC TB426	Graves' disease	Homo sapiens	male	thyroid gland	adult	frozen_sample	long course	Graves' disease	long course	male	TB426.CEL
LC TB442	Graves' disease	Homo sapiens	female	thyroid gland	adult	frozen_sample	long course	Graves' disease	long course	female	TB442.CEL
SC TB373	Graves' disease	Homo sapiens	male	thyroid gland	adult	frozen_sample	short course	Graves' disease	short course	male	TB373E.CEL
SC TB394	Graves' disease	Homo sapiens	female	thyroid gland	adult	frozen_sample	short course	Graves' disease	short course	female	TB394.CEL
SC TB423	Graves' disease	Homo sapiens	female	thyroid gland	adult	frozen_sample	short course	Graves' disease	short course	female	TB423.CEL

Transcription: Gene Expression Atlas



- A curated subset of the ArrayExpress Archive experiments
- Search for gene expression changes under different biological/experimental conditions.

Browse changes in gene expression

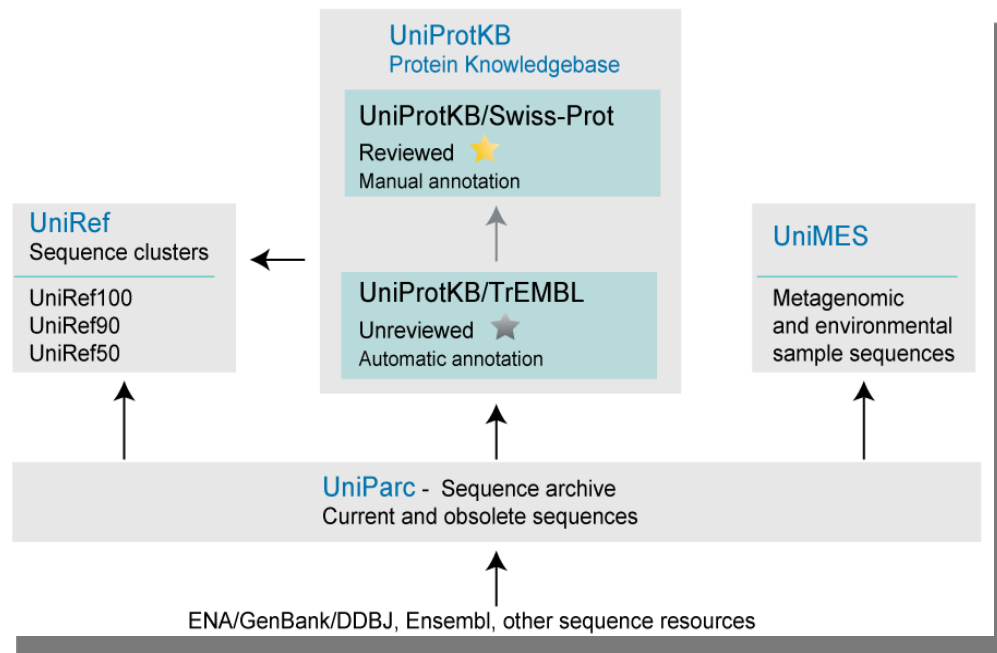
Search by gene or biological condition

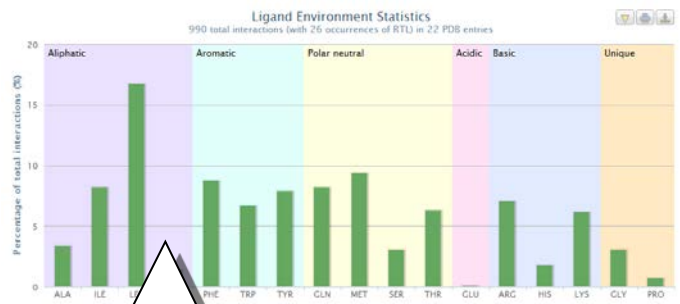
Gene page

Experiment page

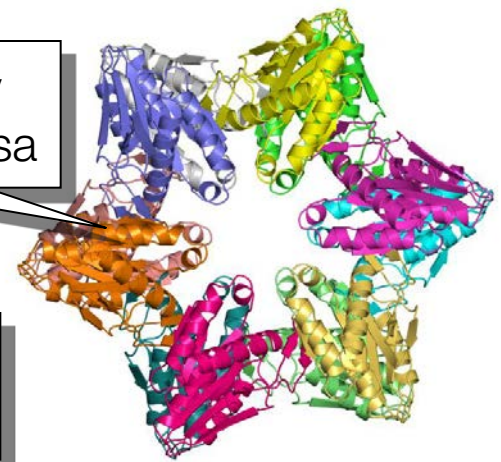
Protein sequences: UniProt

- The world's most comprehensive catalogue of information about proteins
- Richly annotated information on individual proteins and full proteome sets.





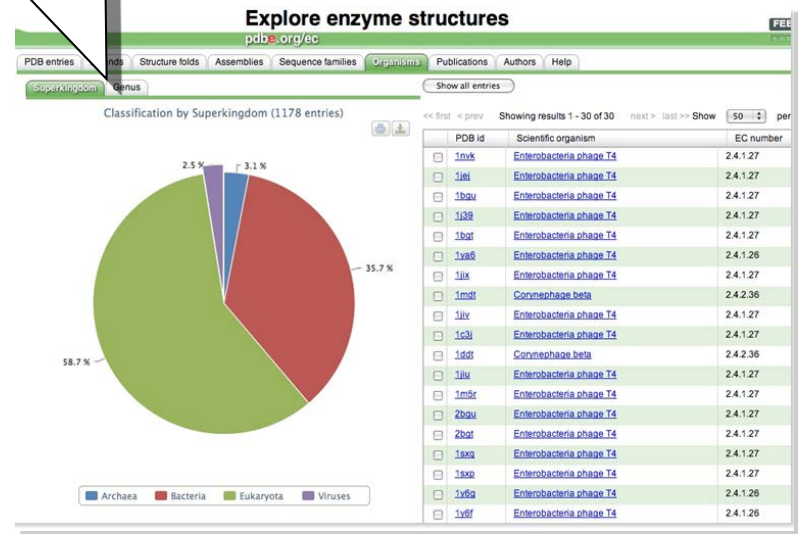
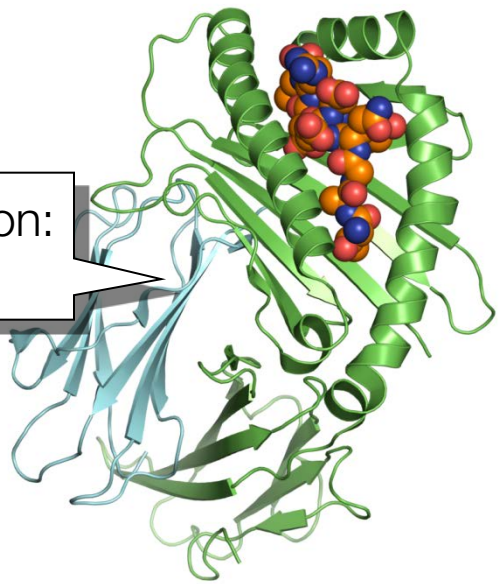
Predict quaternary structure: PDBePisa



Browse structural data: PDBeXplore

Binding sites and motifs: PDBeMotif

Fold comparison: PDBeFold



Chemogenomics: ChEMBL

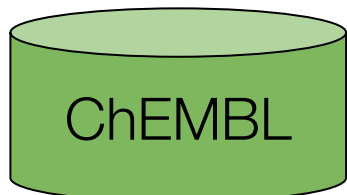
Neglected
Tropical
Disease (NTD)
archive

ChEMBL database

Browse
targets

Target search

Compound
search



Kinase SARfari

Level 3	Level 4	Domain	ChEMBL	Drugs
0	RAF	RAS	2	0
0	ERK5	TEK	10	6
0	RAF	RAF	1	0
0	RAF	RAF	10	3
0	RAF	RAF	2	0
0	RAF	RAF	6	3
0	RAF	RAF	1	1
0	RAF	RAF	1	1
0	RAF	RAF	4	1
0	RAF	RAF	5	1
0	RAF	RAF	2	1
0	RAF	RAF	2	0
0	RAF	RAF	2	1
0	RAF	RAF	3	4
0	RAF	RAF	10	4
0	RAF	RAF	4	2
0	RAF	RAF	4	2

GPCR SARfari

Family	Members	Drugs	
17	Adenosine receptor	10	4
8	Adenosine receptor	1	0
14	Adenosine receptor	13	0
4	Adenosine receptor	2	1

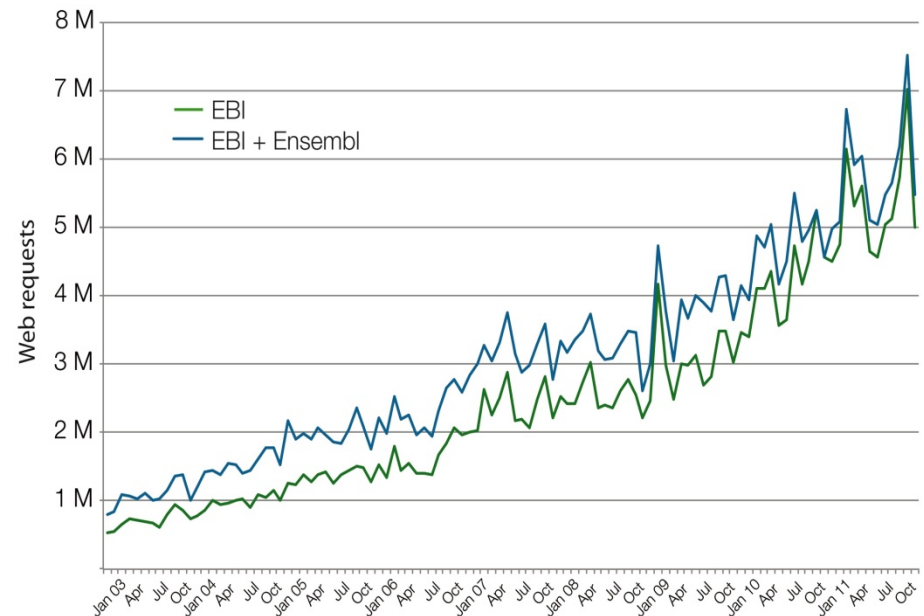


Data management

- Over 5.7 million web requests per day
- Number of unique IPs served grew by 6% between 2010-2011

Data Storage

- Leased two data centres (with €11.4M from UK Research Councils)
- Total disk space: 14 petabytes in 2011.
- Over 800 million cross-references in the databases we serve



Types of Positions in Services at EBI

- Software Engineers
- Database Administrators
- Systems staff
- Curators
- Bioinformaticians
- Project Managers (for internal and external projects)
- Administrators
- User experience



Tools at EMBL-EBI

- Analyse your own data via ~100 tools:
 - Search across our databases for homologous sequences using a suite of tools from classics such as **BLAST** and **FASTA** to recent advances such as **Exonerate** and **PSI-Search**
 - Functionally annotate your sequence with protein domains and important sites using **InterProScan**
 - Align your sequences to discover conserved regions with **Clustal Omega**, **MUSCLE** etc.
 - Perform sequence translation, repeat analysis/masking, trans-membrane topology prediction etc. with **EMBOSS** and other tools
- Results enriched with data from EBI resources

www.ebi.ac.uk/tools

Web Services

Programmatic access available via Web Services

- Remotely run your tasks on our servers + data
- SOAP & REST APIs
- Pre-made clients in Java, PERL, python etc.
- Compatible with workflow tools such as Galaxy, YABI, Taverna
- Tools, data retrieval, even EBI Search
- Same programs, data and results enrichment as running via the web pages

www.ebi.ac.uk/tools/webservices

Research

www.ebi.ac.uk/groups



Key facts about research at EMBL-EBI

- A unique environment for bioinformatics research
- Nine dedicated research groups
- Seven services teams also carry out R&D
- Research and services are mutually supportive



Diversity in discipline: first degree

Genomes



Ewan
Birney



Paul
Flicek



Nick
Goldman



Oliver
Stegle



Transcriptomes



Alvis
Brazma



Anton
Enright



John
Marioni



Key



biology/medicine



chemistry/chem
engineering



maths



physics

Proteins



Janet
Thornton



Gerard
Kleywegt



Christoph
Steinbeck



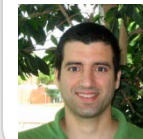
John
Overington



Pathways and systems



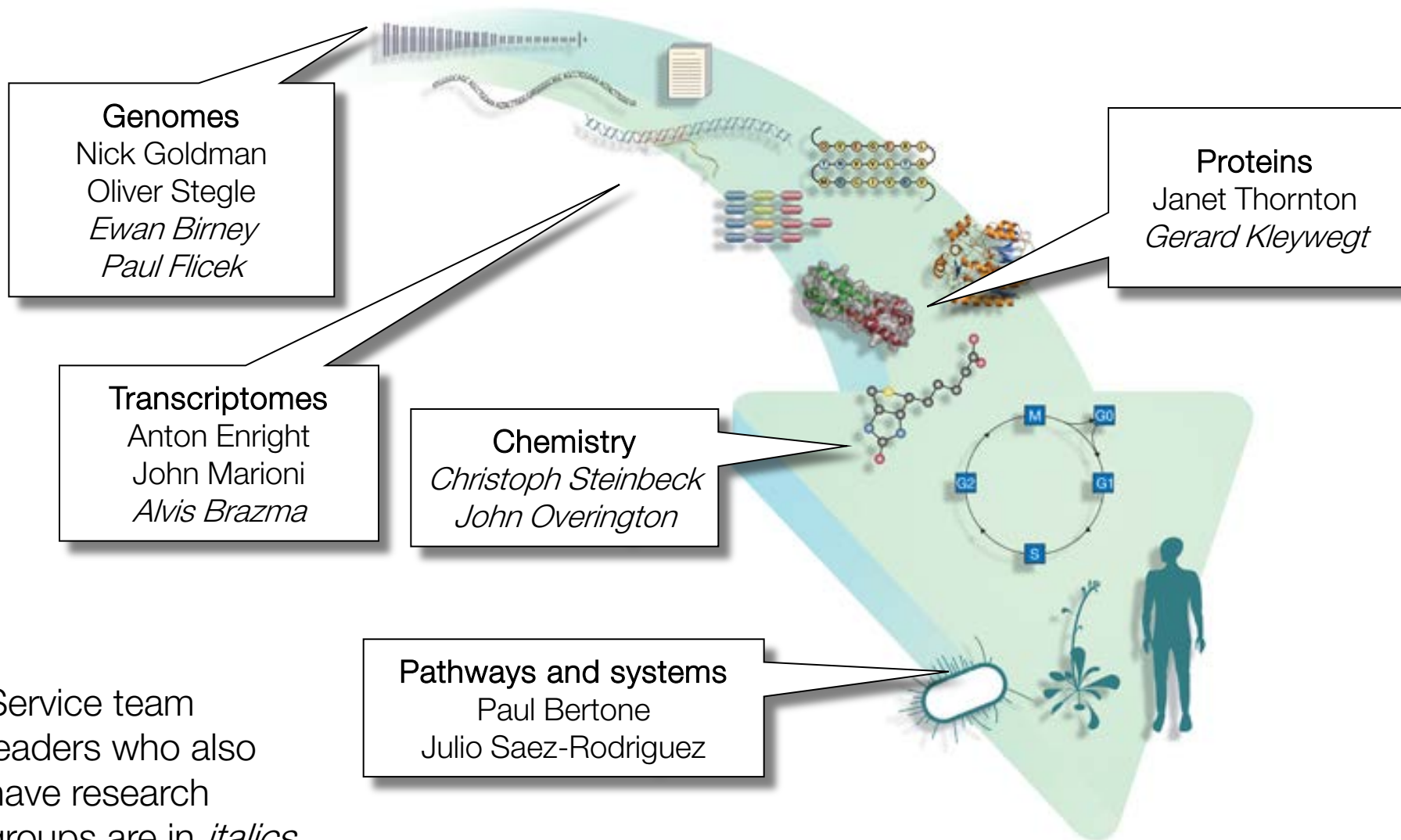
Paul
Bertone



Julio Saez-
Rodriguez



Research themes

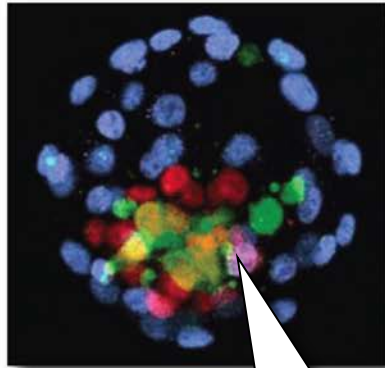


Examples of EMBL-EBI research

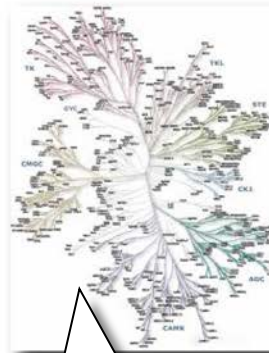
What is the molecular basis of ageing?



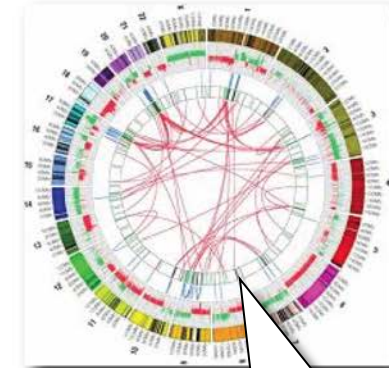
How do the neurons of someone with Parkinson's disease signal differently from healthy neurons?



What makes a stem cell decide to become skin or muscle?



Which of these proteins will make good targets for drugs?



Which of these changes to a genome's structure drive cancer?

New research group leaders – to join in 2013

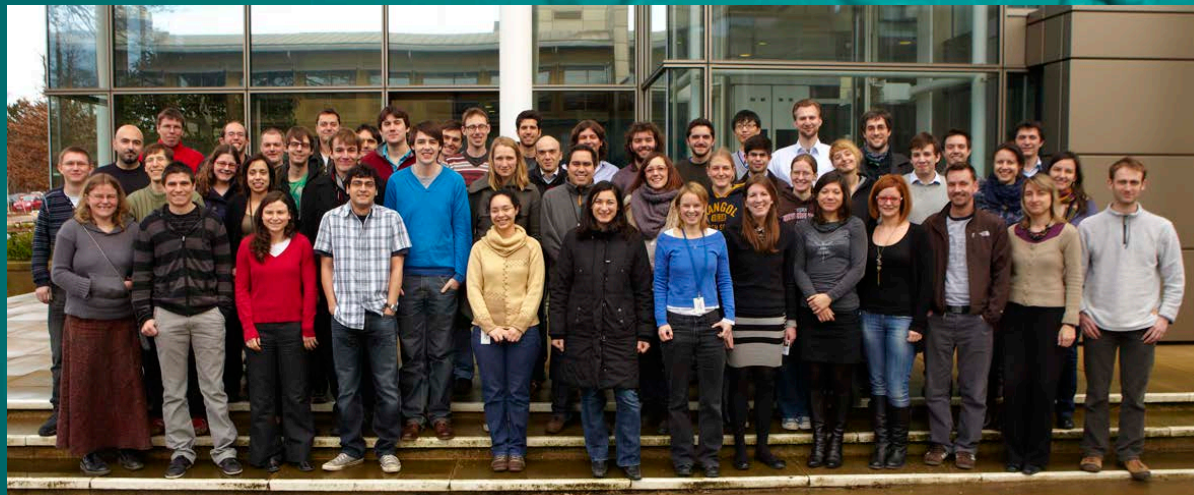
● Sarah Teichmann

● Pedro Beltrao



Training

www.ebi.ac.uk/training

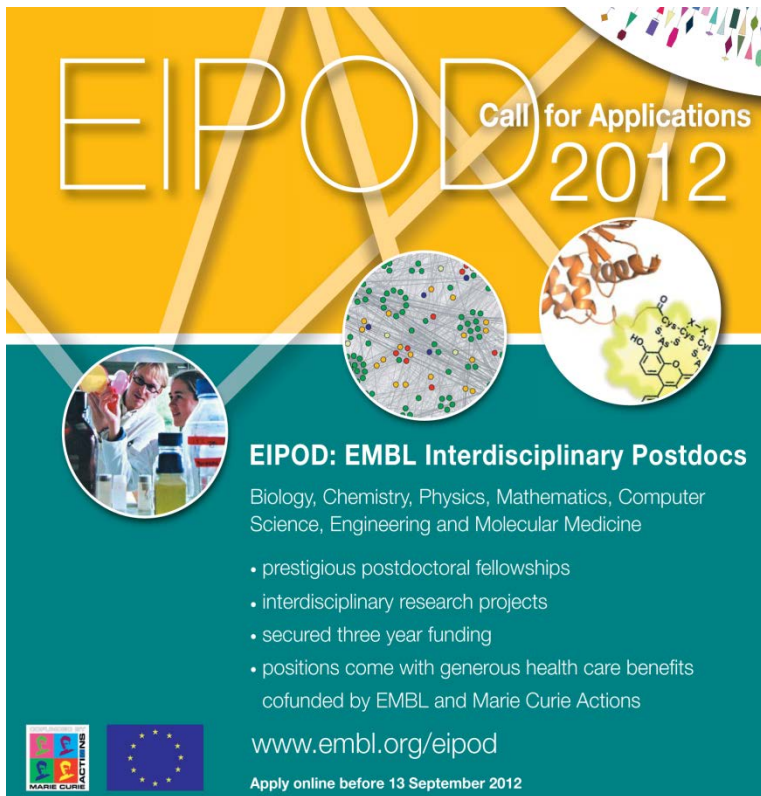


EMBL International PhD Programme 2012



- All applications must be made online at:
www.embl.org/phdprogramme
- The 2013 Fall selection will open around mid-April
- Contact:
predocs@embl.de

Postdocs at EMBL-EBI



EIPOD Call for Applications 2012

EIPOD: EMBL Interdisciplinary Postdocs
Biology, Chemistry, Physics, Mathematics, Computer Science, Engineering and Molecular Medicine

- prestigious postdoctoral fellowships
- interdisciplinary research projects
- secured three year funding
- positions come with generous health care benefits cofunded by EMBL and Marie Curie Actions

www.embl.org/eipod
Apply online before 13 September 2012

Logos for Marie Curie Actions and the European Union are visible at the bottom left.

EIPOD: EMBL-sponsored interdisciplinary fellowships



ESPOD
Call for applications: 2012

EBI-Sanger Postdoctoral Programme

- Deadline for applications: 15 August 2012
- Combine experimental and computational approaches to address important challenges in life science research
- Propose your own project or select from pre-defined projects:
 - Systematic characterisation of Sanger human iPSCs (SH-iPSCs)
 - Modelling the genetic-epigenetic regulatory pathways in hematopoietic cells
 - Genes, pathways, modules and organs that regulate bone mass
 - Comparative genomic insights into parasite genome function
 - Adaptive evolution of olfactory systems in mammals

www.ebi.ac.uk/training/postdoc/ESPOD

Logos for Wellcome Trust Sanger Institute and EMBL-EBI are visible at the bottom right.

ESPOD: EBI-Sanger combined experimental and computational fellowships

User training

www.ebi.ac.uk/training

Learn to make the most of Europe's most widely used databases for life scientists



Train at EMBL-EBI

Gain hands-on experience in our state-of-the-art training room



Train at your place

Choose the training modules relevant to you and your colleagues and let us bring the Bioinformatics Roadshow* to you



Train online

Beta release launched: train wherever and whenever you like and at your own pace

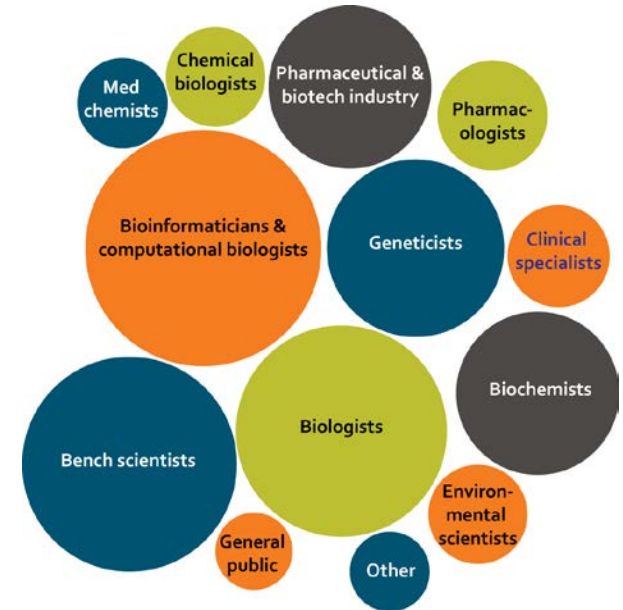
* The Bioinformatics Roadshow is funded by the European Commission under SLING, grant agreement number 226073 (Integrating Activity) within the Research Infrastructures of the FP7 Capacities Specific Programme.

EMBL-EBI



Why is user training important?

- Growing and diversifying user community
- ~3 million life scientists in Europe, each a potential bioinformatics user
- Data analysis is the major bottleneck in research
- Many molecular life scientists feel under-qualified to use publicly available data.



EMTRAIN

- A sustainable, pan-European platform for education and training
- Covers the whole life-cycle of medicines research:
 - basic science
 - clinical development
 - pharmacovigilance
- On-course: Europe's first biomedical post-graduate course catalogue
- Quality criteria listed by course providers.



EMTRAIN European Medicines Research Training Network

Join us on LinkedIn

LinkedIn



We are on-course®
on-course
inspiring science

www.emtrain.eu

Interactions with industry

www.ebi.ac.uk/industry

The EBI Industry Programme

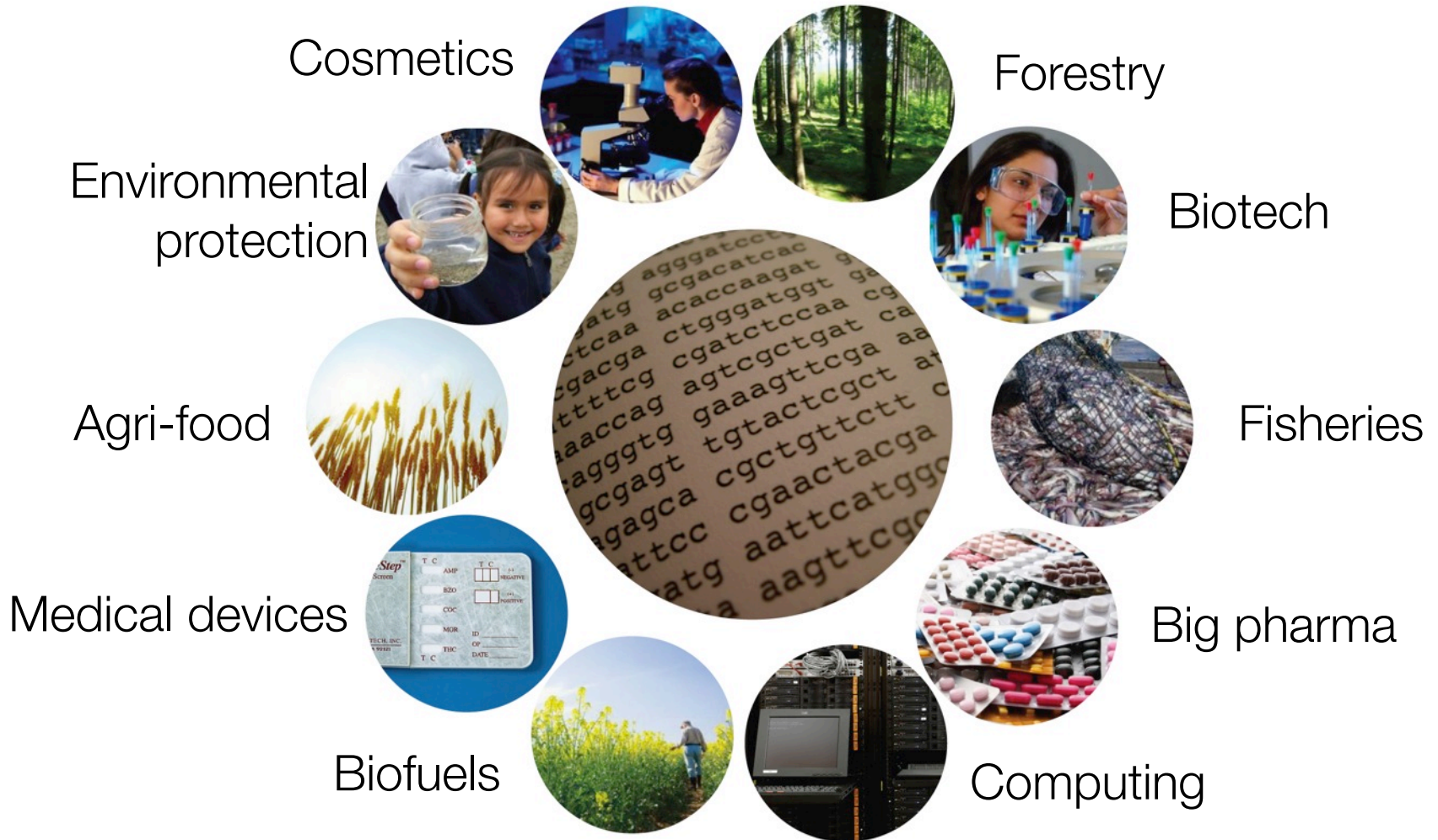
- Large companies join the EBI Industry Programme which provides
 - Expert training
 - Standards development
 - Technical development
 - Precompetitive research collaboration
 - Networking opportunities
- This fosters collaborations between industry and EBI
- It enables industry to adapt quickly to – and maximise the benefit from – innovations in bioinformatics.
- It enables EBI to develop services which fulfil the needs of industry



“The Industry Programme’s regular meetings foster inter-company interactions as we collaborate on special projects and liaise on other industry initiatives”

-- *Bertram Weiss, Principal Scientist,
Target Discovery, Bayer Schering
Pharma*

Supporting the bioindustries



Industry Programme members

- Astellas Pharma Inc.
- AstraZeneca
- Bayer Pharma AG
- Boehringer Ingelheim
- Eli Lilly and Company
- F. Hoffmann-La Roche
- Galderma
- GlaxoSmithKline
- Johnson & Johnson
- Pharmaceutical R&D Recherche & Développement
- Merck Serono S.A.
- Nestlé Research Centre
- Novartis Pharma AG
- Novo Nordisk
- Orion Pharma
- Pfizer Ltd
- Syngenta
- Sanofi-Aventis
- UCB
- Unilever

Small and medium-sized enterprises

- SMEs are major drivers of the economy
- Turnover is high: needs are often more short-term
- EMBL-EBI data and services can add value to business processes immediately



Pictured: attendees of the 2011 SME Forum in Piemonte, Italy

Coordinating bioinformatics in Europe

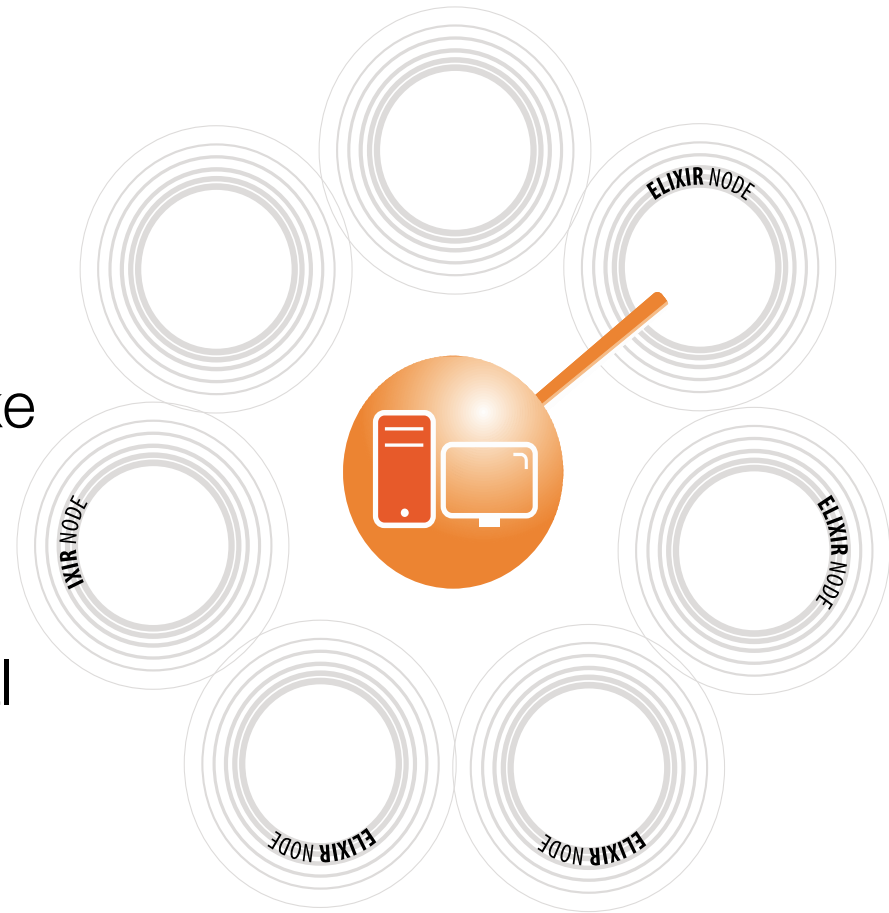
ELIXIR

- EMBL-EBI is a major driver in ELIXIR, the pan-European research infrastructure for biological information
- The goal of ELIXIR:
 - Build a sustainable European infrastructure for biological information
 - Support life science research and its translation to medicine, the environment, the bioindustries and society.

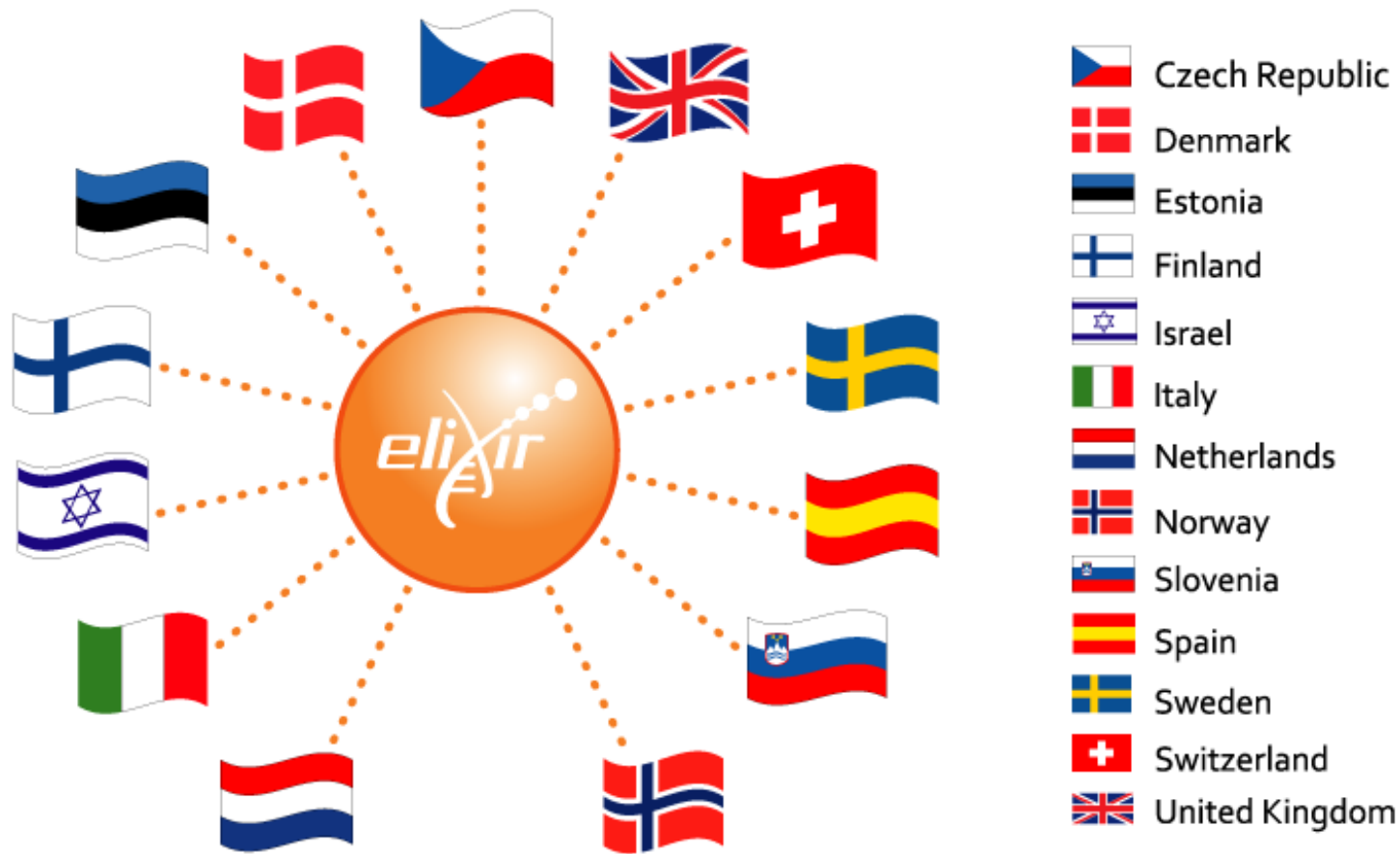


A distributed infrastructure

- Central hub at EMBL-EBI
- Hub provides technical coordination
- Nodes in member states make scientific and technical contributions
- Collaboration through bilateral service level agreements between hub and nodes.



Signatories to the ELIXIR Memorandum of Understanding



Thank you! Any questions?

