

# EMBL-EBI workshop: Data and tools for transcriptomics and protein biology

[https://www.ebi.ac.uk/~lauhuema/  
workshop/fg\\_Malta2018](https://www.ebi.ac.uk/~lauhuema/workshop/fg_Malta2018)

Laura Huerta, PhD  
Senior Scientific Curator  
[lauhuema@ebi.ac.uk](mailto:lauhuema@ebi.ac.uk)  
Malta, 31 July 2018





UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# *A bit about your speaker...*

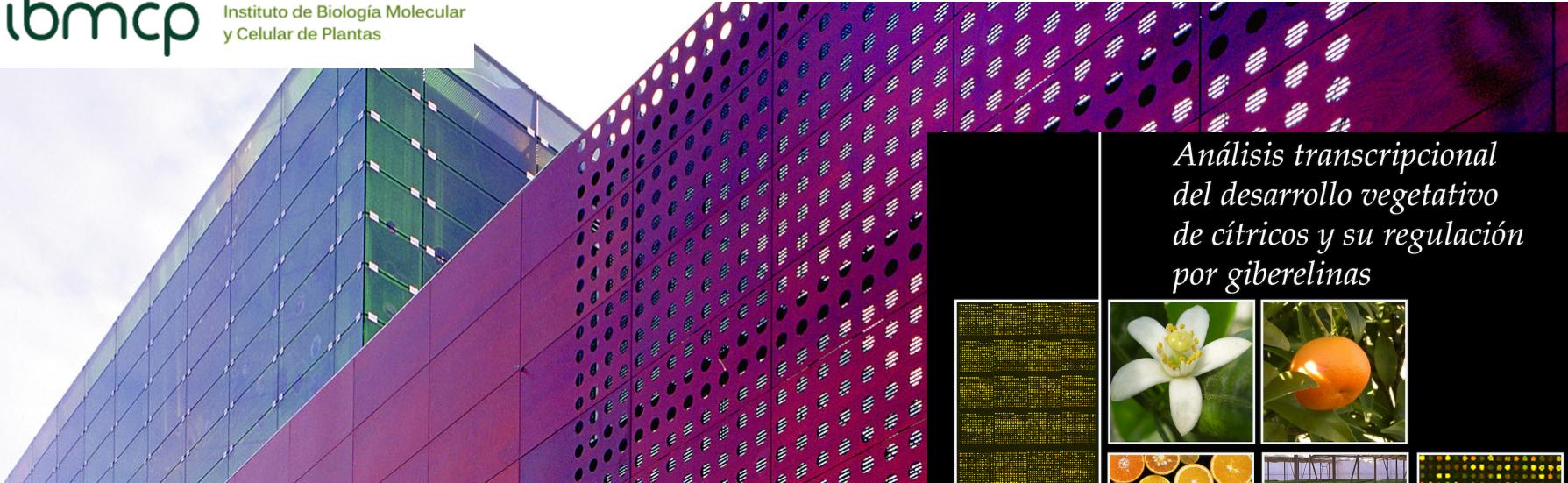


- Agronomic Engineer, major in Biotechnology
- Cloning and characterisation of a GA 20-oxidase gene from *Citrus sinensis* (L.) Osbeck. Prof. José Luis García Martínez (CSIC)

# *A bit about your speaker...*

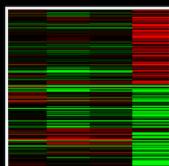
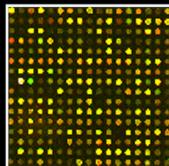


Instituto de Biología Molecular  
y Celular de Plantas



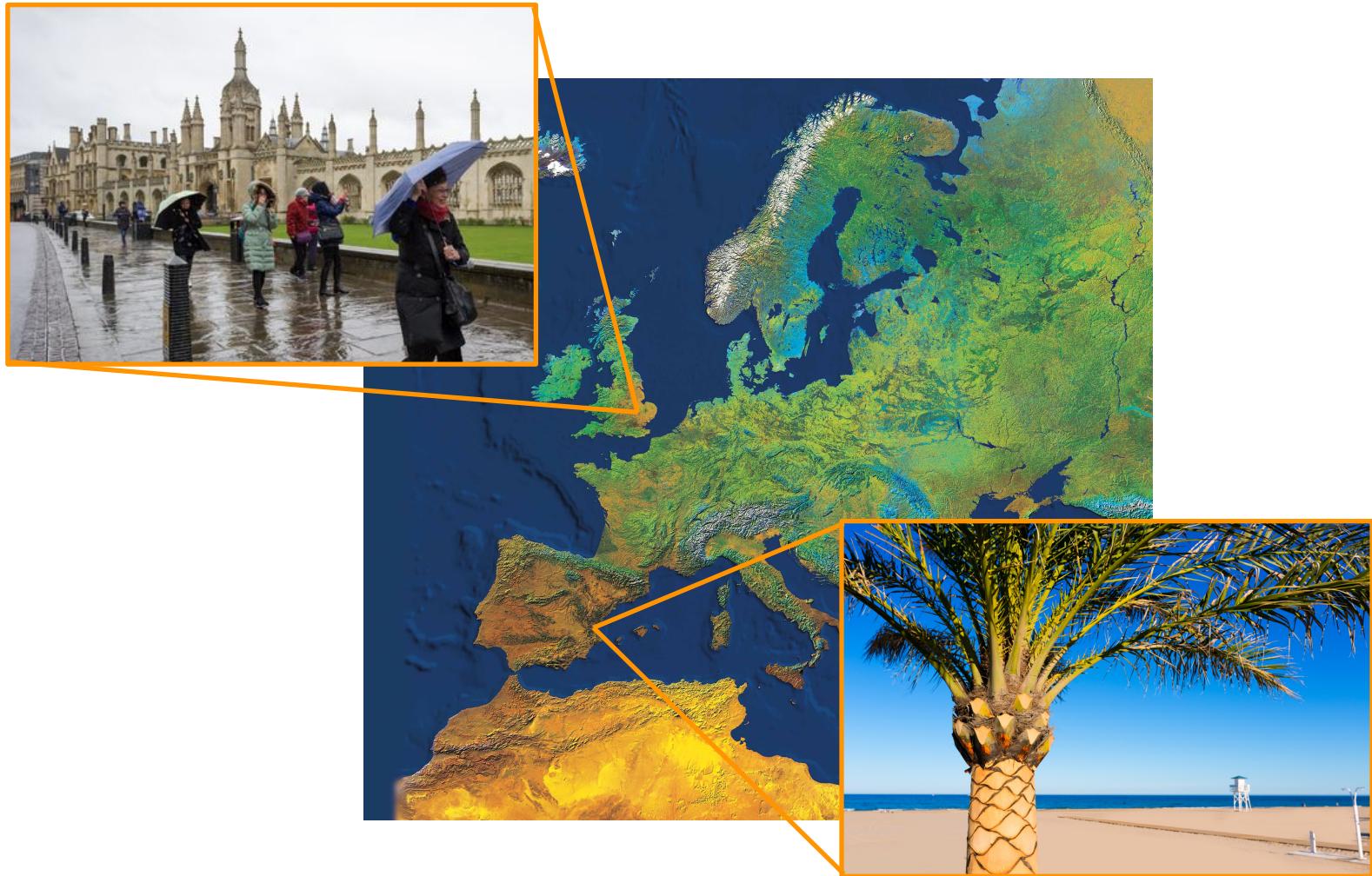
<http://www.ibmcp.csic.es>

*Análisis transcripcional  
del desarrollo vegetativo  
de cítricos y su regulación  
por giberelinas*



TESIS DOCTORAL  
*Laura Huerta Martínez*

# *A bit about your speaker...*

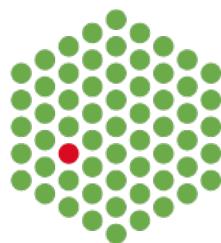




WELLCOME  
GENOME  
CAMPUS



# EMBL-EBI



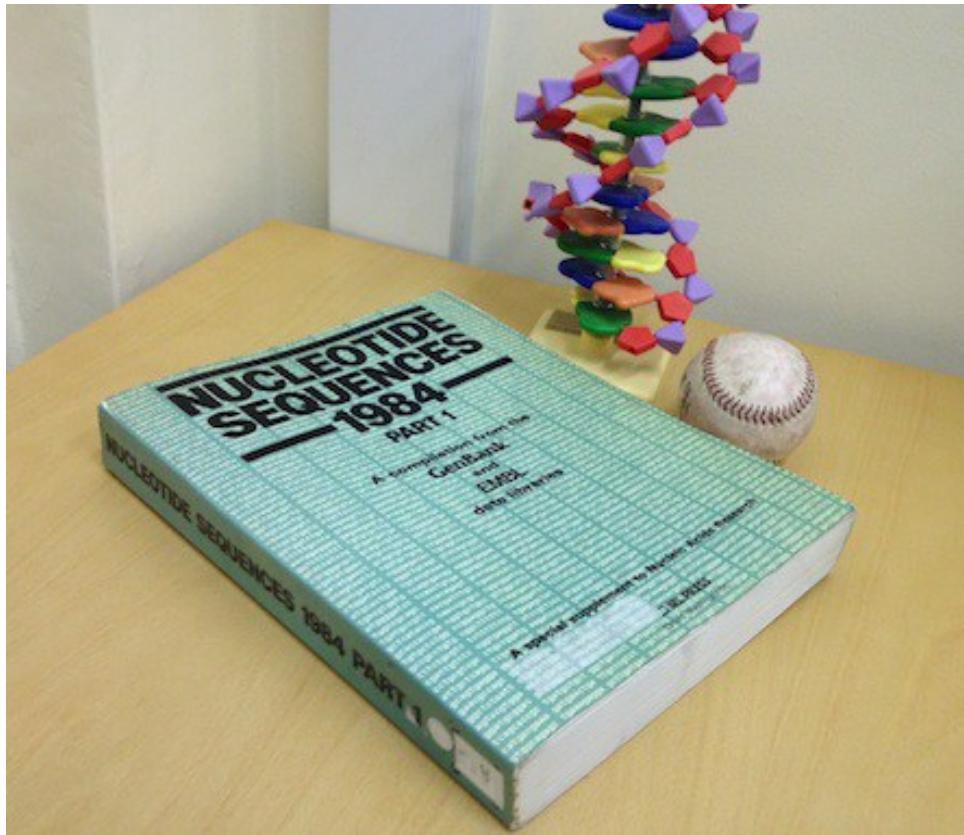
# EMBL-EBI workshop: Data and tools for transcriptomics and protein biology

## Introduction to EMBL-EBI resources

Laura Huerta, PhD  
Senior Scientific Curator  
[lauhuema@ebi.ac.uk](mailto:lauhuema@ebi.ac.uk)  
Malta, 31 July 2018



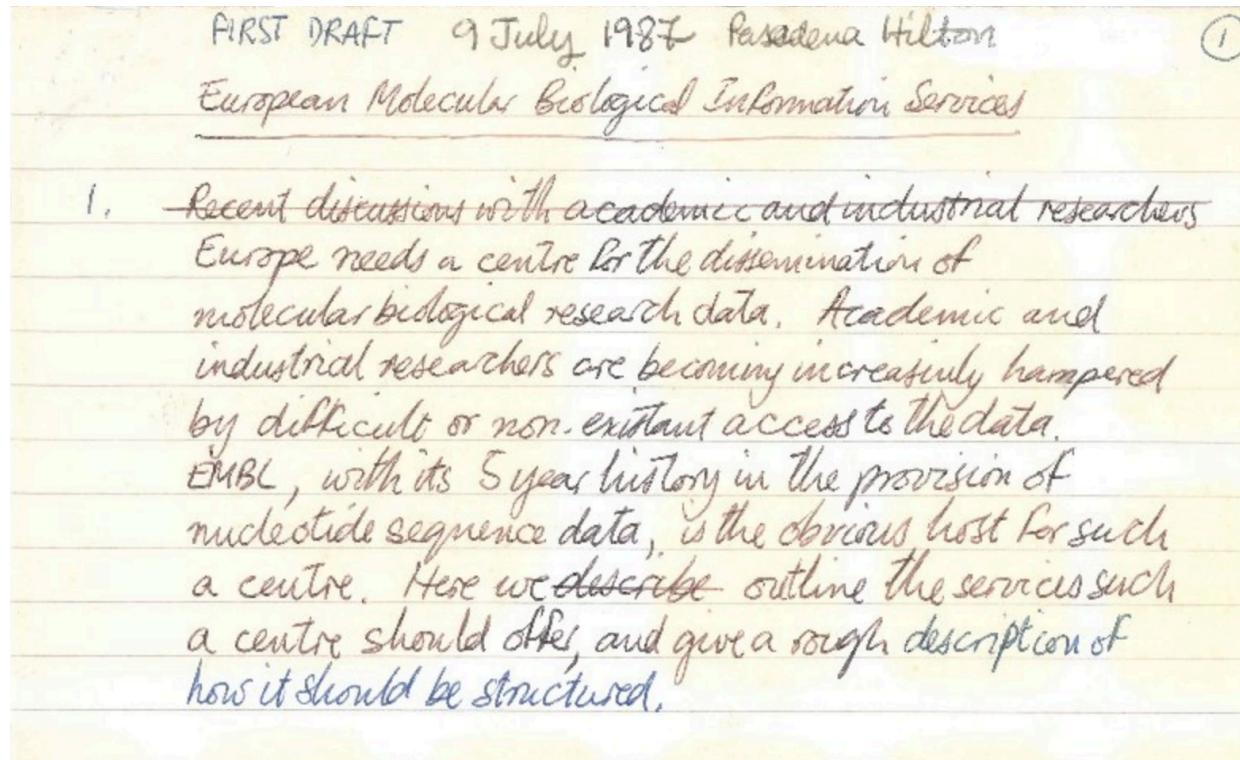
# Life science: a revolution in biological data



Nucleotide sequences 1984 Part 1 A compilation  
from the GenBank and EMBL data libraries

# Life science: a revolution in biological data

The first draft of Graham Cameron's proposal for a "European Molecular Biological Information Services".



# Life science: a revolution in biological data



# The European Bioinformatics Institute (EBI)

The home for big data in biology



To provide freely available data and bioinformatics services to the scientific community

[www.ebi.ac.uk/](http://www.ebi.ac.uk/)

# What is EMBL-EBI?

- Europe's home for biological data services, research and training
- A trusted data provider for the life sciences:
  - 150 Petabytes of storage
  - > 40,000 CPU Cores
- Part of the European Molecular Biology Laboratory, an intergovernmental research organisation
- International: 600 members of staff from 60 nations
- Home of the ELIXIR Technical hub.

# The European Molecular Biology Laboratory

+60 nationalities

Heidelberg, Germany



Main Laboratory

>1800 personnel

Hinxton, Cambridge, UK



Bioinformatics

6 sites in Europe

Grenoble, France



Structural Biology



Tissue Biology, Disease Modeling  
Barcelona, Spain



Mouse Biology  
Monterotondo, Rome, Italy



Structural Biology  
Hamburg, Germany

# EMBL member states



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

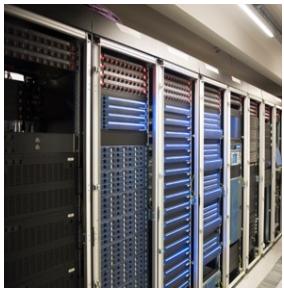
Associate member states: Argentina, Australia

Prospect member states: Lithuania, Poland



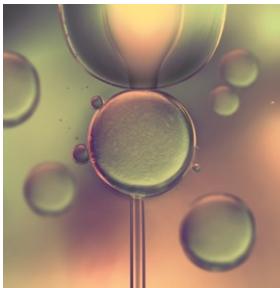
# What is EMBL-EBI? – Our mission

## Services



Deliver scientific services

## Research



Deliver excellent research

## Training



Train the next generation of scientists

## Industry



Engage with European industry

## ELIXIR



Coordinate bioinformatics and data in Europe

# Bioinformatics services

Data and tools to support life science  
research

[www.ebi.ac.uk/services](http://www.ebi.ac.uk/services)

# What services do we provide?

...provide tools to help researchers use it

Labs around the world send us their data and we...

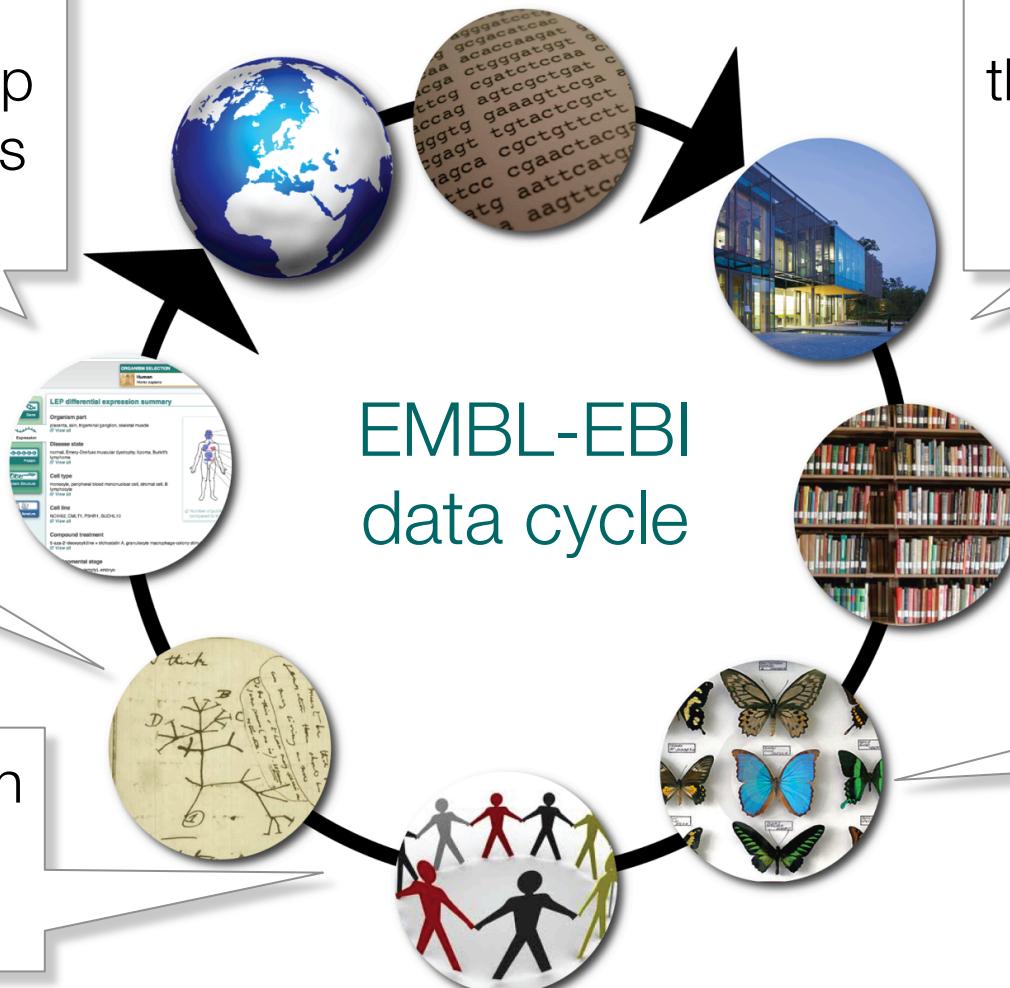
## EMBL-EBI data cycle

Analyse it

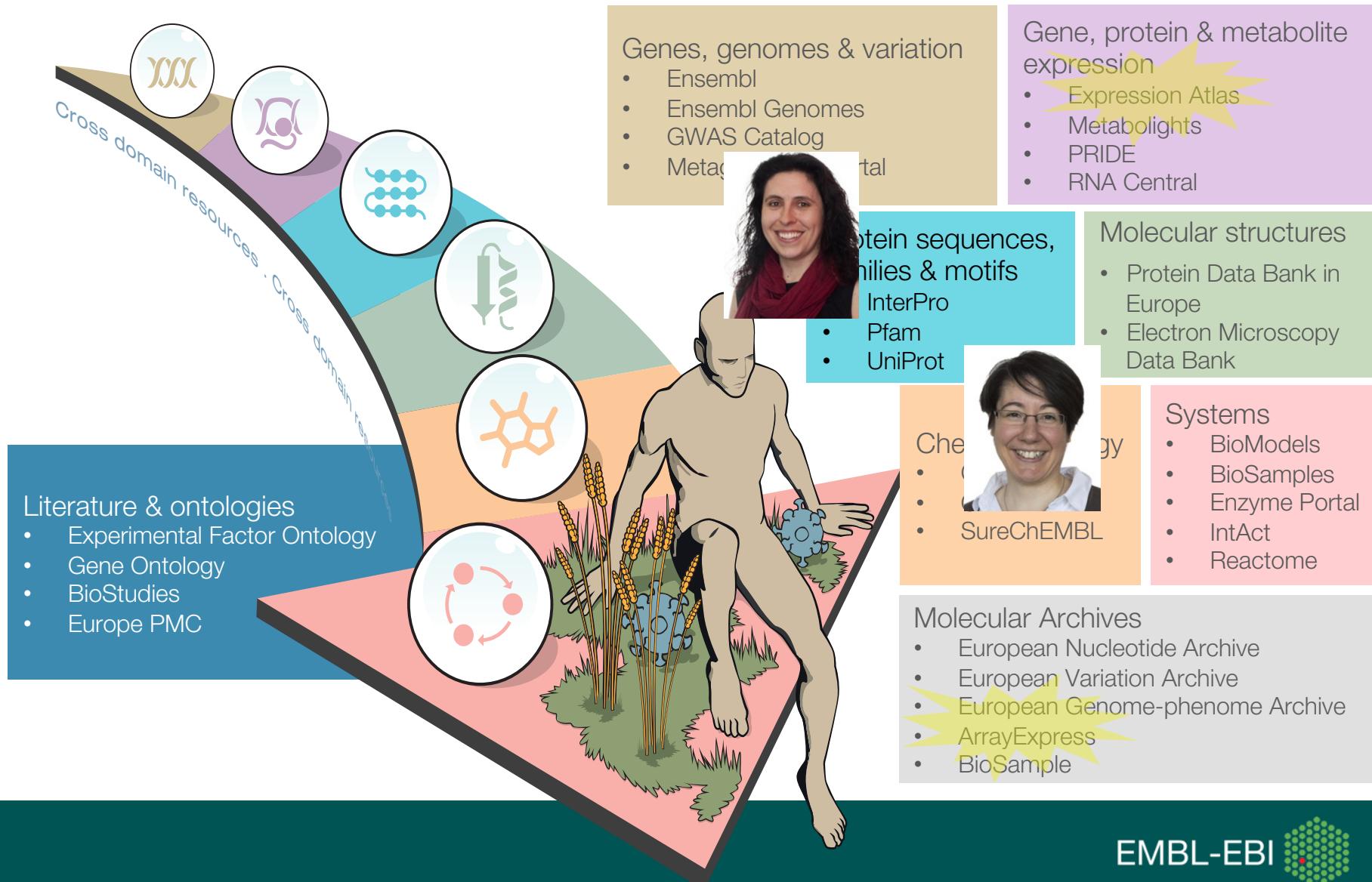
Share it with other data providers

Archive it

Classify it



# Data resources at EMBL-EBI



# “Essential to research”

**70%**

of users agreed EMBL-EBI services are “essential to their research”\*

\*From a survey of over 1800 users in 2013 and 2014

**Top 1**

for the processes of generation and scholarly communication of scientific knowledge\*

*\*From the Ranking Web of Research Infrastructures*

Europe PMC offers

**4.4 million**

full-text articles, 1.2 million of which are Open Access

Over

**350 000**

scientists used EMBL-EBI’s online training in 2016

# Saving cost, enabling innovation

**79%**

“EMBL-EBI services reduced the cost of exploiting relevant data.”

**93%**

“EMBL-EBI services reduced the time needed to find required data.”

**79%**

“EMBL-EBI services permitted us to explore more novel research questions.”

**81 %**

“EMBL-EBI services reduced duplication of effort.”

# Research

Leading computational biology  
research since 1994

[www.ebi.ac.uk/research](http://www.ebi.ac.uk/research)

# Research groups at EMBL-EBI



Alex Bateman    Ewan Birney    Pedro Beltrao    Alvis Brazma    Rob Finn    Paul Flicek    Andrew Leach    Moritz Gerstung



Nick Goldman    Zamin Iqbal    John Marioni    Evangelia Petsalaki    Oliver Stegle    Janet Thornton    Virginie Uhlmann    Daniel Zerbino

# Postdoctoral fellowships at EMBL-EBI

## Interdisciplinary

EIPOD projects bring together scientific fields that are usually separate, or transfer techniques to a novel context.

## Genome Campus

ESPODs: joint Sanger–EMBL-EBI fellowships combine experimental and computational approaches.

## Biomedicine

EBPODs are joint projects with EMBL-EBI, NIHR Cambridge Biomedical Research Centre and the University of Cambridge School of the Biological Sciences.

# Training

Overview

Train at EMBL-EBI

Train outside EMBL-EBI

Train online

Webinars

About

10 Years

Contact

# Bioinformatics training

EMBL-EBI Training

## For scientists working at all levels

We train scientists at all levels to get the most out of publicly available biological data.

Follow us on Twitter and sign up to the EMBL courses and conferences newsletter.

[www.ebi.ac.uk/training](http://www.ebi.ac.uk/training)

Sponsoring 10 years of public training → Topic filtered by Gene expression [X]

Filter by: Topic

- biomodels
- DNA & RNA
- genomes
- Gene expression
- Proteins
- Structures
- Systems
- Ontologies
- Chemical biology
- Literature
- Cross domain

Filter by: Type

Train at EMBL-EBI

Train outside EMBL-EBI

## Training events

May 2018

24

EMBL-EBI Workshop: Tools for functional genomics data

Conrad Hotel, Seoul, South Korea - 10 Gukjegeumyung-ro, Yeoeuido-dong, Yeongdeungpo-gu, Seoul, Seoul, South Korea

24th May

This course will provide participants with an introduction to functional genomics resources at EMBL-EBI, in particular to ArrayExpress and the submission tool Annotare.

## Online courses

Start now

[ArrayExpress: Discover functional genomics data quickly and easily](#)

Author(s): Anja Füllgrabe

ArrayExpress is a database of functional genomics data. This course will give you an overview of how these data are stored in ArrayExpress and will teach you how to effectively search and retrieve data from the ArrayExpress website . Course...

Start now

[ArrayExpress: Quick tour](#)

Author(s): Melissa Burke



Celebrating 10 Years of Training

# Building capacity: bioinformatics training

## Train online

We provide free access to materials from EMBL-EBI courses, allowing you to choose when, where and how you learn

## Webinars

Our collection of free seminars focuses on our range of data resources and new features and tools to boost your research

## Courses at EMBL-EBI

Our on-site training courses help life-science researchers in all domains to make the most of public data resources

## Courses at host sites

We send our expert trainers to host organisations throughout the world, where they provide training on bioinformatics

# Industry engagement

Support and collaboration

[www.ebi.ac.uk/industry](http://www.ebi.ac.uk/industry)

# EMBL-EBI industry programme members

## Pharmaceutical

**abbvie**



**Celgene**

**Lilly**

**NOVARTIS**

**SANOFI**

**astellas**  
Leading Light for Life

**Biogen**

**genzyme**

**MedImmune**



**astex<sup>®</sup>**  
pharmaceuticals

**Boehringer  
Ingelheim**

**gsk**

**MERCK**

**Pfizer**

**ucb**

**AstraZeneca**

**Bristol-Myers Squibb**

**janssen**  
PHARMACEUTICAL COMPANIES  
of Johnson & Johnson

**MSD**  
*Be well*

**Roche**

## Agri-food and consumer goods

**Bayer CropScience**

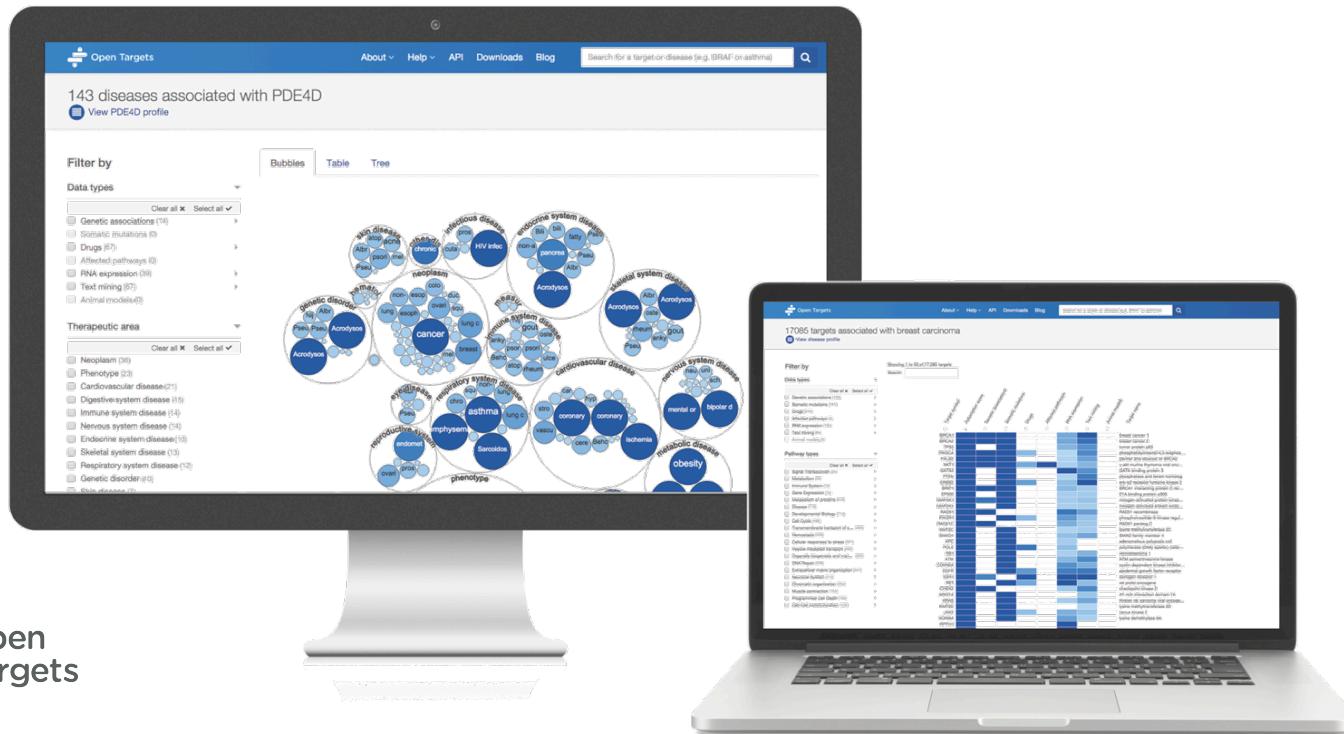
**syngenta**

**Unilever**



# Open Targets: Target Validation Platform

A partnership to transform drug discovery through the systematic identification and prioritisation of targets



# Elixir

Supporting the coordination of  
biological data provision in Europe

[www.elixir-europe.org](http://www.elixir-europe.org)



ELIXIR unites Europe's leading life science organisations in managing and safeguarding the data generated by publicly funded research.

It coordinates, integrates and sustains data resources across its member states, enabling scientists to access services that are vital for their research.



# Navigating EBI – www.ebi.ac.uk

The home for big data in biology

Open data  
EMBL-EBI shares data from life science experiments.

More about EMBL-EBI's impact in our annual report ➤  
Data from 2016

Our unique Search service helps you explore dozens of biological data resources.  
More about EBI Search ➤

All Find a gene, protein or chemical

Example searches: blast keratin bf1

Find a tool for your data analysis.

Find a tool ➤

Share your scientific data with the world.

Deposit data ➤

## We are EMBL-EBI

The European Bioinformatics Institute (EMBL-EBI) is part of EMBL, Europe's flagship laboratory for the life sciences. More about EMBL-EBI and our impact ➤

## Training

Access a wealth of world-leading training in bioinformatics and scientific service provision, regardless of your career stage or sector ➤

## Data resources

Explore our open data resources to enrich your research. Browse data, perform analyses or share your own results. ➤

## Industry

Explore our knowledge-exchange Industry Programme and take part in translational partnerships and projects ➤

## Research

Find out about our research groups, postdoctoral schemes and PhD Programme ➤

## ELIXIR

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

# Navigating EBI: tools & data resources

## Services

Overview

A to Z

Data submission

Support

The European Bioinformatics Institute (EMBL-EBI) maintains the world's most comprehensive range of freely available and up-to-date molecular data resources.

Developed in collaboration with our colleagues worldwide, our services let you share data, perform complex queries and analyse the results in different ways. You can work locally by downloading our data and software, or use our [web services](#) to access our resources programmatically.

— You can read more about our services in the journal [Nucleic Acids Research](#)

## Tools & Data Resources

### Tools

#### Clustal Omega



Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.

[Multiple sequence alignment](#)

### Data resources

#### Ensembl

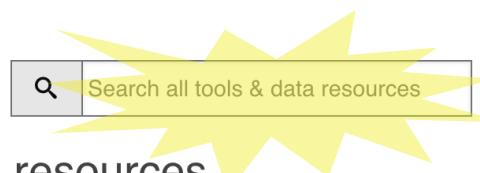


Genome browser, API and database, providing access to reference genome annotation

#### UniProt



A comprehensive resource for protein sequence and functional annotation.



### Browse by type

XXX DNA & RNA	Gene Expression	Proteins
Structures	Systems	Chemical biology
Ontologies	Literature	Cross domain

# Navigating EBI: tools & data resources

Overview

A to Z

Data submission

Support

Proteins | [View all tools and data resources](#) ›

## Tools & Data Resources

### Tools

#### BLAST [protein]



Fast local similarity search tool for protein sequence databases.

[Sequence similarity search](#)

#### ChEMBL BLAST Search

BLAST search against protein targets stored in ChEMBL database

[Sequence similarity search](#)

#### Clustal Omega



Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.

### Data resources

#### Complex Portal



A manually curated, encyclopaedic resource of macromolecular complexes from model organisms.

#### Enzyme Portal



Integrated enzyme data from EMBL-EBI resources.

#### IntAct



A curated database of molecular interactions.

#### IntEnz



Reference enzyme nomenclature and function.



Filter these tools & data resources

### Browse by type

DNA & RNA	Gene Expression	Proteins
Structures	Systems	Chemical biology
Ontologies	Literature	Cross domain

### Programmatic access

EMBL-EBI web services allow you to query our large biological data resources programmatically, so that you can develop data analysis pipelines or integrate public data with your own applications. The Web Services technology we use are built on open standards to ensure client and server software from various sources will work well together.

[Browse EMBL-EBI web services](#)

# Navigating EBI: data submission

The home for big data in biology

Open data  
EMBL-EBI shares data from life science experiments.

More about EMBL-EBI's impact in our annual report ➤  
Data from 2016

Our unique Search service helps you explore dozens of biological data resources.  
More about EBI Search ➤

All ▾ Find a gene, protein or chemical

Example searches: blast keratin bf1

Find a tool for your data analysis.  
 Find a tool ➤

Share your scientific data with the world.  
 Deposit data ➤

## We are EMBL-EBI

The European Bioinformatics Institute (EMBL-EBI) is part of EMBL, Europe's flagship laboratory for the life sciences. More about EMBL-EBI and our impact. ➤

## Training

Access a wealth of world-leading training in bioinformatics and scientific service provision, regardless of your career stage or sector ➤

## Data resources

Explore our open data resources to enrich your research. Browse data, perform analyses or share your own results. ➤

## Industry

Explore our knowledge-exchange Industry Programme and take part in translational partnerships and projects ➤

## Research

Find out about our research groups, postdoctoral schemes and PhD Programme ➤

## ELIXIR

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

# Navigating EBI: data submission

## Services

Overview    A to Z    Data submission    Support

SERVICES / DATA SUBMISSION

## Data submission

Use this data submission wizard to find the right archive for your data in a few simple steps.

1 What **type of data** do you have?

DNA/RNA sequence

Expression data

Protein data

Structures

Systems

Chemical biology

Ontologies

Multi-omics or other cross-domain study

## Why submit data to an archive?

Submission of primary data and derived information to public data repositories is an essential step in the scientific process. Through submission, the scientific community is fed the raw materials for the building and maintenance of the complete and up-to-date data sets that support searches and analysis on the latest sequences, structures and molecular profiles of living systems. Serving as a complement to the literature publication process and supporting early data sharing, the EBI offers a number of submission services appropriate for different types and scales of data.

## All EMBL-EBI data repositories

[Array Express](#) ➤ functional genomics data

[BioModels](#) ➤ computational models

[BioSamples](#) ➤ reference sample data

[ChEBI](#) ➤ chemical entities

[DGVA](#) ➤ structural genetic variation data

[EFO](#) ➤ experimental variables

[EGA](#) ➤ human data that requires controlled access

[EMPIAR](#) ➤ raw image data

[ENA](#) ➤ nucleotide sequence data

[EVA](#) ➤ genetic variation data

[GO](#) ➤ Gene Ontology annotations

[IntAct](#) ➤ molecular interactions

[IntEnz](#) ➤ enzyme nomenclature

# Navigating EBI: EBI search

The home for big data in biology

Open data  
EMBL-EBI shares data from life science experiments.

More about EMBL-EBI's impact in our annual report ➤  
Data from 2016

Our unique Search service helps you explore dozens of biological data resources.  
More about EBI Search ➤

All ▾ BRCA1

Example searches: blast keratin bf1

Find a tool for your data analysis.

Share your scientific data with the world.

Deposit data ➤

## We are EMBL-EBI

The European Bioinformatics Institute (EMBL-EBI) is part of EMBL, Europe's flagship laboratory for the life sciences. More about EMBL-EBI and our impact ➤

## Training

Access a wealth of world-leading training in bioinformatics and scientific service provision, regardless of your career stage or sector ➤

## Data resources

Explore our open data resources to enrich your research. Browse data, perform analyses or share your own results. ➤

## Industry

Explore our knowledge-exchange Industry Programme and take part in translational partnerships and projects ➤

## Research

Find out about our research groups, postdoctoral schemes and PhD Programme ➤

## ELIXIR

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

BRCA1

X



Examples: VAV\_HUMAN , tp53 , Sulston ...

Build Query

Help &amp; Documentation

About EBI Search

Feedback

## Search results for

**BRCA1**

Showing 45 results out of 72,953 in All results

### Filter your results

#### Source

**All results** (72,953)

Genomes &amp; metagenomes (9,507)

Nucleotide sequences (15,735)

Protein sequences (21,181)

Macromolecular structures (86)

Bioactive molecules (93)

Gene expression (8,432)

Diseases (180)

Molecular interactions (444)

Reactions &amp; pathways (653)

Protein families (68)

Protein expression data (16)

Enzymes (7)

Literature (12,767)

Samples &amp; ontologies (3,766)

EMBL-EBI web (18)

#### Gene & protein summaries (includes expression, structures, literature...) (3 results)

 **BRCA1**, DNA repair associated  
**BRCA1** (RNF53, undefined)  
Human(Homo sapiens) **Breast cancer 1, early onset**  
**Brca1** (ENSMUSG00000017146, undefined)  
House Mouse(Mus musculus)[More...](#)

#### Protein sequences (21,181 results)

**P38398** (BRCA1\_HUMAN) Source: UniProtKB (ID: BRCA1\_HUMAN)

Breast cancer type 1 susceptibility protein

Homo sapiens (Reviewed)

Secondary accession number(s): E9PFZ0, O15129, Q1RMC1, Q3LRJ0, Q3LRJ6, Q6IN79, Q7KYU9

# Navigating EBI: EBI search

## Gene & protein summary for BRCA1

ORGANISMS

human  
*Homo sapiens*

Gene Expression Protein Protein Structure Literature

 View in Ensembl

### BRCA1, DNA repair associated

#### Gene Information and Sequence

- BRCA1 spans 125951 bps of chromosome 17 from 43044295 to 43170245.
- BRCA1 has 33 transcripts containing a total of 110 exons on the reverse strand.
- Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see [article](#).
- [View the gene sequence in Ensembl](#).
- [View the chromosome region for this gene in Ensembl](#).

#### Variations

- BRCA1 has 33271 SNPs.
- [View sequence variations such as polymorphisms, along with genotypes and disease associations in Ensembl](#).

#### Orthologues

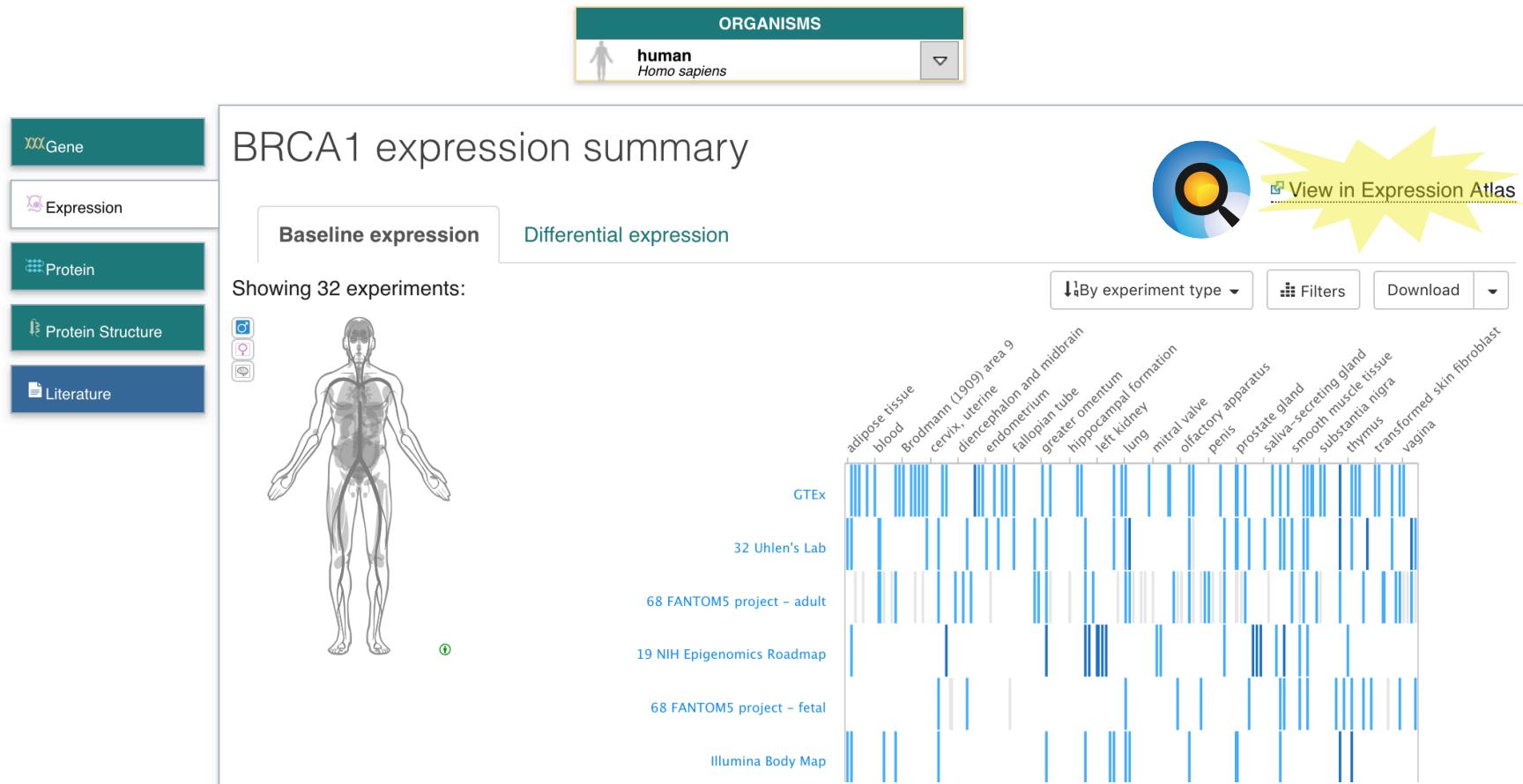
- BRCA1 has 101 orthologues in Ensembl
- [View homology between species inferred from a gene tree in Ensembl](#).

#### Paralogues

- BRCA1 has no paralogues in Ensembl

# Navigating EBI: EBI search

## Gene & protein summary for BRCA1



# Navigating EBI: EBI search

## Gene & protein summary for BRCA1

ORGANISMS

human  
*Homo sapiens*

Breast cancer type 1 susceptibility protein

**Functions**

E3 ubiquitin-protein ligase that specifically mediates the formation of 'Lys-6'-linked polyubiquitin chains and plays a central role in DNA repair by facilitating cellular responses to DNA damage. It is unclear whether it also mediates the formation of other ubiquitin-protein ligase activity is required for its tumor suppressor function. The BRCA1-BAR of cellular pathways such as DNA damage repair, ubiquitination and transcriptional regulation to maintain genomic stability. Regulates centrosomal microtubule nucleation. Required for normal cell cycle progression from G2 to mitosis. Required for appropriate cell cycle arrests after ionizing irradiation in both the S-phase and the G2 phase of the cell cycle. Involved in transcriptional regulation of P21 in response to DNA damage. Required for FANCD2 targeting to sites of DNA damage. May function as a transcriptional regulator. Inhibits lipid synthesis by binding to inactive phosphorylated ACACA and preventing its dephosphorylation. Contributes to homologous recombination repair (HRR) via its direct interaction with PALB2, fine-tunes recombinational repair partly through its modulatory role in the PALB2-dependent loading of BRCA2-RAD51 repair machinery at DNA breaks. Component of the BRCA1-RBBP8 complex which regulates CHEK1 activation and controls cell cycle G2/M checkpoints on DNA damage via BRCA1-mediated ubiquitination of RBBP8. Acts as a transcriptional activator (PubMed:20160719).

[View annotation in UniProt](#)

 View in UniProt

 InterPro E3 range

InterPro domain architecture

# Navigating EBI: EBI search

## Gene & protein summary for BRCA1

ORGANISMS

human  
*Homo sapiens*

▼

Gene Expression Protein Protein Structure Literature

23 protein structures available

Structure of human BRCA1 BRCT in complex with BAAT peptide

Method

X-ray diffraction

Experiment

Resolution: 2.2Å

R-Factor: 20.021%

Free R-Factor: 26.063%

Dates

Deposited: 14-12-2012

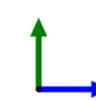
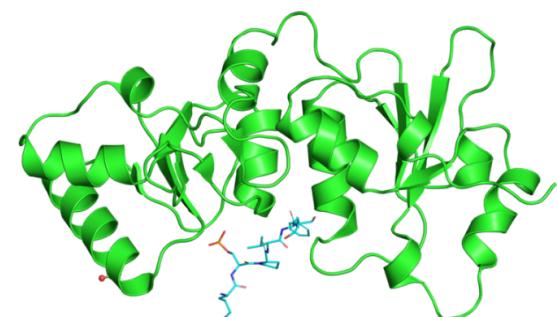
Released: 30-10-2013

Revised: 25-12-2013

Deposited by

Liu, X.,Ladias, J.A.A.

[View in PDBe](#)



# Navigating EBI: EBI search

## Gene & protein summary for BRCA1

ORGANISMS

human  
*Homo sapiens*

All Articles (36178)

New developments in the genetic diagnosis of short stature.  
(PMID:29787394)  
Jee YH, Baron J, Nilsson O.  
*Curr Opin Pediatr*[2018, 30(4):541-547]  
Cited:0 times

Engagement with Genetic Information and Uptake of Genetic Testing: the Role of Trust and Personal Cancer History.  
(PMID:28105554)  
Roberts MC, Taber JM, Klein WM.  
*J Cancer Educ*[2018, 33(4):893-900]  
Cited:1 times

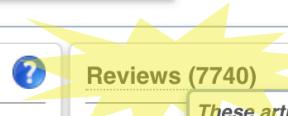
Overcoming Chemoresistance in Cancer via Combined MicroRNA Therapeutics with Anti-Cancer Drugs using Multifunctional Magnetic Core-Shell Nanoparticles.  
(PMID:30028120)  
Yin PT, Pongkulapa T, Cho HY, Han J, Pasquale N, Rabie H,

Reviews (7740)

Therapeutic cancer.  
(PMID:30023007 PMCID:PMC6047242)  
Shi Y, Jin C, et al.  
*Mol Cancer*[2018, 17(1):100]  
Cited:0 times

Dancing with the DNA damage response: next-generation anti-cancer therapeutic strategies.  
(PMID:30023007 PMCID:PMC6047242)  
Minchom A, Aversa C, Lopez J.  
*Ther Adv Med Oncol*[2018, 10:1758835918786658]  
Cited:0 times

Role of RNF20 in cancer development and progression - a comprehensive review.  
(PMID:29934362 PMCID:PMC6043722)  
Sethi G, Shanmugam MK, Arfuso F, Kumar AP.  
*Biosci Rep*[2018, 38(4)]  
Cited:0 times



### Europe PMC

# EBI is like a treasure trove of data



Janet Thornton, Director Emeritus, European Bioinformatics Institute

EMBL-EBI

# EMBL-EBI workshop: Data and tools for transcriptomics and protein biology

## Introduction to EMBL-EBI resources

Laura Huerta, PhD  
Senior Scientific Curator  
[lauhuema@ebi.ac.uk](mailto:lauhuema@ebi.ac.uk)  
Malta, 31 July 2018

