

Unlocking the potential of big data in biology: Applications of bioinformatics

Anna Swan

Senior Scientific Training Officer - Digital Learning

Wellcome Connecting Science - Genome Academy, 11 April 2024

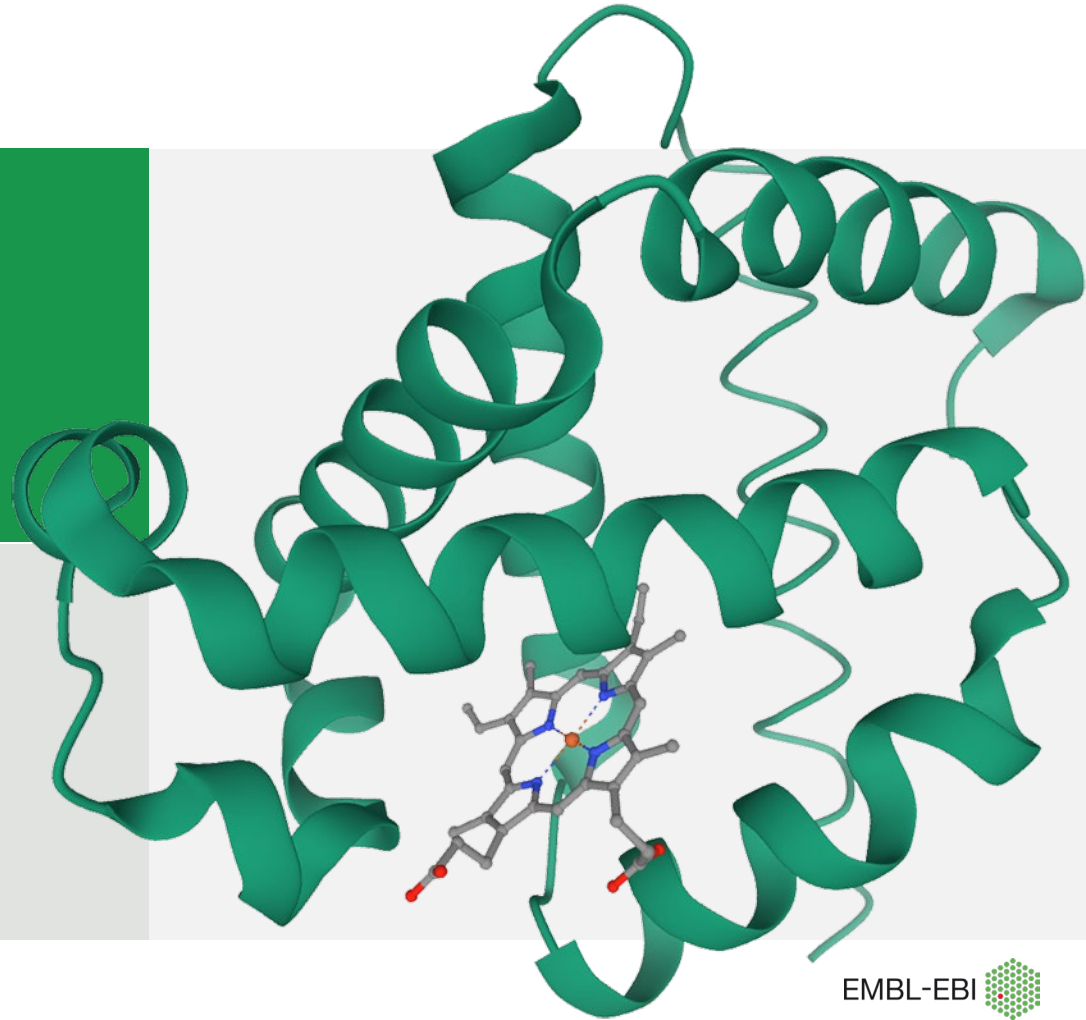
bit.ly/GA-bioinformatics-2024





Bioinformatics

- Bioinformatics is the study of biology (bio) with computers (informatics).
- Involves storing, managing and analysing huge datasets.





What is EMBL-EBI?

- The home of big data in biology
- One of the six sites of the European Molecular Biology Laboratory (EMBL)
- Intergovernmental organisation

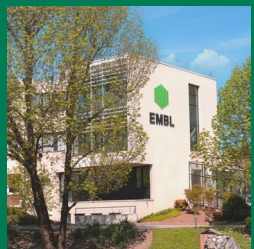


The European Molecular Biology Laboratory



EMBL-EBI

Bioinformatics



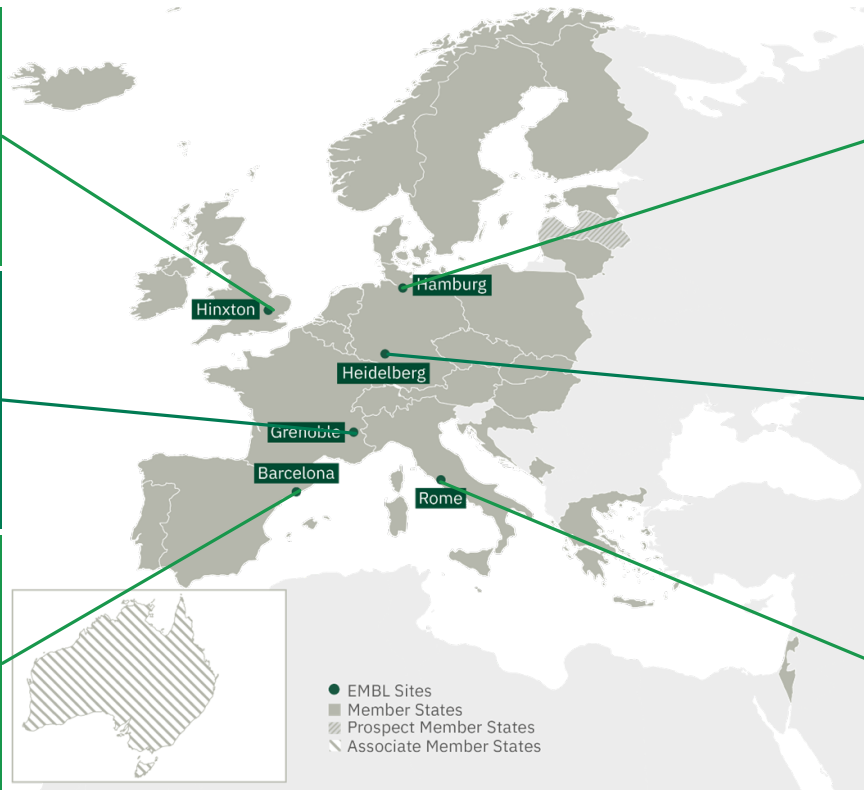
Grenoble

Structural biology



Barcelona

Tissue biology
and disease
modelling



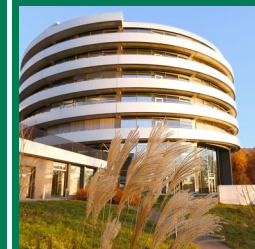
Hamburg

Structural biology



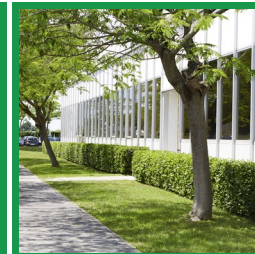
Heidelberg

Life sciences



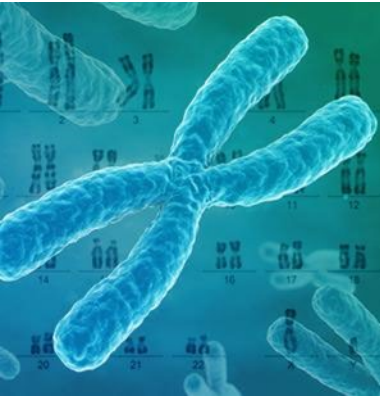
Rome

Epigenetics
and neurobiology





What does EMBL-EBI do?



Provide data resources for the life sciences

Perform excellent research

Train the next generation of scientists

Work with the private sector

Coordinate bioinformatics in Europe

Data resources at EMBL-EBI

The image displays a composite of three web interfaces. On the left is the Ensembl genome browser for Human (GRCh38.p14), showing the BRCA2 gene region (13 32,315,086-32,400,268) with various tracks for gene-based displays, annotations, and variants. In the center is the UniProt database entry for A2VEY9 - APP_DROME, detailing its function as a palmitoyltransferase app, subcellular location, and other protein characteristics. On the right is the AlphaFold Protein Structure Database homepage, which features a search bar and lists examples of protein structures like 'Free fatty acid receptor 2' and 'Q5VSL9' from E. coli. The AlphaFold page also notes it was developed by Google DeepMind and EMBL-EBI.

AlphaFold DB provides open access to over 200 million protein structure predictions to accelerate scientific research.

Data resources at EMBL-EBI



Chemicals, molecules and drug discovery

ChEBI
ChEMBL
MetaboLights
Open Targets
SureChEMBL



Genes, genomes and RNA

ArrayExpress
Ensembl
European Nucleotide Archive
Expression Atlas
HGNC
MGnify
Rfam
RNACentral
VectorBase
WormBase



Proteins

AlphaFold DB
Enzyme Portal
InterPro
PDBe
Pfam
PRIDE
UniProt



Imaging and cellular structure

BioImage Archive
Electron Microscopy Data Bank
EMPIAR



Genetic variation and disease data

COVID-19 Data Platform
European Genome-phenome Archive
European Variation Archive
Mouse informatics

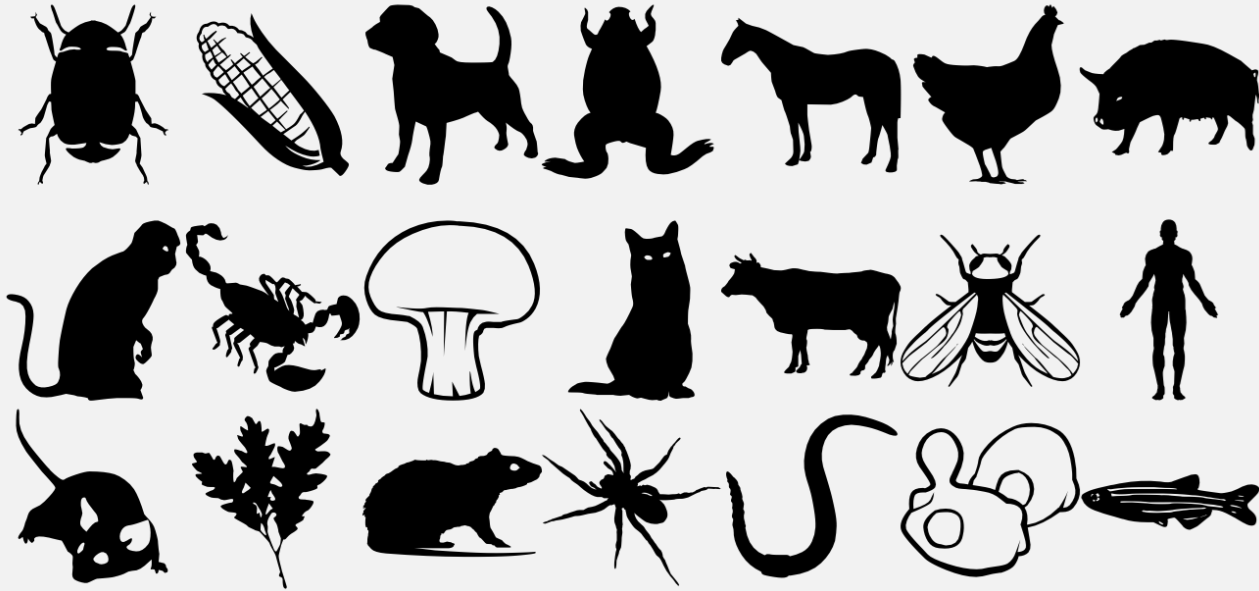


Literature and knowledge management

BioModels
BioSamples
BioStudies
Complex Portal
Europe PMC
GWAS Catalog
IntAct
OmicsDI
Ontologies
Reactome



Data for many species





Searching for a gene, protein or chemical

EMBL's European Bioinformatics Institute
EMBL-EBI
Unleashing the potential of big data in biology

Find a gene, protein or chemical All
Example searches: blast keratin bfl1 | About EBI Search

EMBL's European Bioinformatics Institute
EBI Search
Access all EMBL-EBI resources

Examples: VAV_HUMAN, tp53, Sulston... [Advanced search](#)

Search results for *mstn*
Showing **41** results out of **8,548** in All results
[Give us feedback on these results](#)

[Find data resources](#) [Submit data](#)

Latest news

Funding announcement
UKRI funding supports EMBL-EBI infrastructure upgrade in response to increasing data demand
30 Nov 2023

The Global Biodivers enabling biodiversity worldwide
29 Nov 2023

Filter your results

Source
All results (8,548)
Genomes & metagenomes (369)
Nucleotide sequences (3,205)
Protein sequences (1,674)
Macromolecular structures (667)
Bioactive molecules (2)
Gene expression (181)
Diseases (24)
Molecular interactions (12)
Gene-Disease Associations (592)
Reactions & pathways (12)
Protein families (45)
Protein expression data (2)
Enzymes (1)
Literature (1,475)
Samples & ontologies (287)

Gene and protein summaries (2 results • includes expression, structures, literature...)

Myostatin
MSTN (ENSG00000138379)
Human (Homo sapiens)

Myostatin
Mstn (ENSMUSG00000026100)
House Mouse (Mus musculus)

Samples & ontologies (287 results)
Source: Taxonomy (ID: 1219472)
Reporter vector pGL3-MSTN-3.8kb

Data resources at EMBL-EBI

107 million requests to our data resources on an average day

1 Scientists generate data, make discoveries

2 Deposit with EMBL-EBI on publication

3 We archive and share data with global collaborators and all scientists



4 We classify, enrich, combine and analyse

5 We distribute both raw and “value added” data resources

6 Scientists design new experiments on basis of shared global knowledge

What is open data?

- Open data can be freely used, re-used and redistributed by anyone.
- When research data is open others can use it to ask new questions and get new insights.
- Open data saves repeating experiments
- Open data drives new discoveries.
- EMBL-EBI data resources are open data.





We don't wear lab coats

Biologists, physicists
mathematicians

Software engineers

Biocurators



Bioinformaticians

Data wranglers

And more!

What skills are needed for bioinformatics?

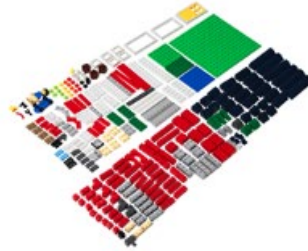
1 Data Collection



2 Data Preparation



3 Data Visualization



4 Data Analysis



5 Data Storytelling



www.effective-datastorytelling.com

Organisational

Problem-solving

Creativity and Experimentation

Communication



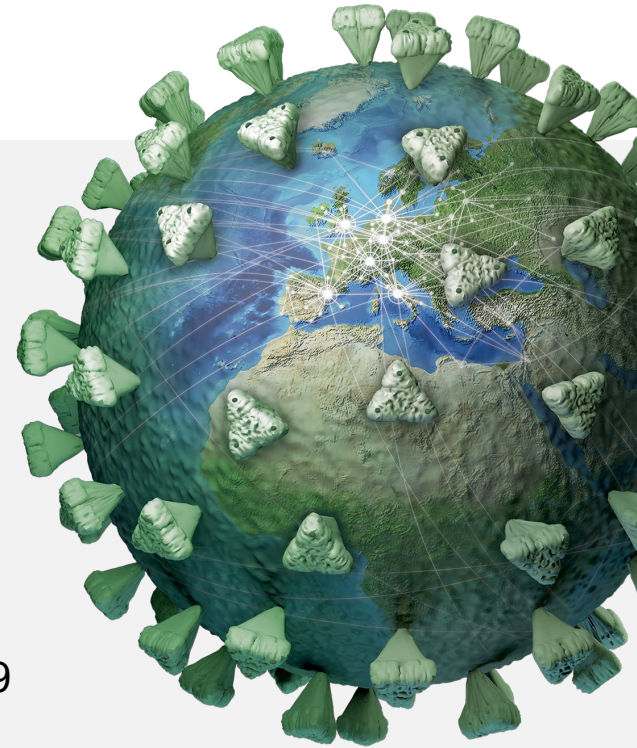
Pandemic preparedness

Data science was essential in the COVID-19 pandemic.

EMBL-EBI supported the pandemic response:

- Set up the COVID-19 Data Portal to access SARS-CoV-2 molecular and genomic data from all over the world
- Supported countries to set up data sharing platforms
- Revealed insights on new ‘variants’
- Analysed molecular causes of different immune responses
- Identified existing drugs that could be used to treat COVID-19

EMBL-EBI and collaborators are helping to improve European pandemic preparedness.





Sustainable food production

Bioinformatics helps to feed a growing population in a changing climate.

- Plant genomics – identify which species will be most tolerant to drought and pests while still providing optimum nutrition
- Pests and pollinators – genomics can inform strategies for dealing with pests while protecting pollinators
- Precision breeding – linking genes to traits, farmers and breeders can make food production more sustainable



ACGACTG
TCTTAGTAA
GCTGTAGGCAC
CTCCAATTGTG



Biodiversity conservation

Bioinformatics helps us

- understand and protect biodiversity
- develop clean technologies to reduce environmental pollution

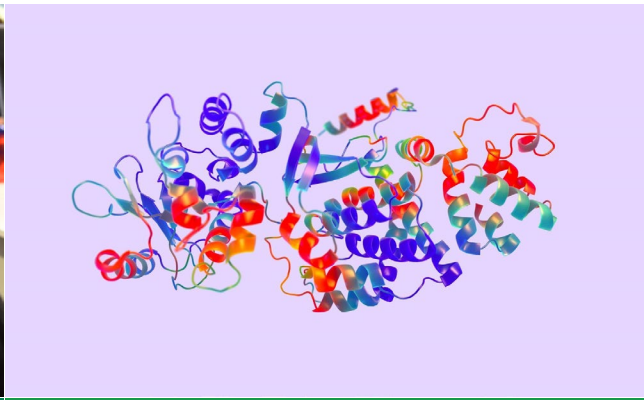
Darwin Tree of Life project

- Sequence 77,000 eukaryotic species in Britain and Ireland





A growing field



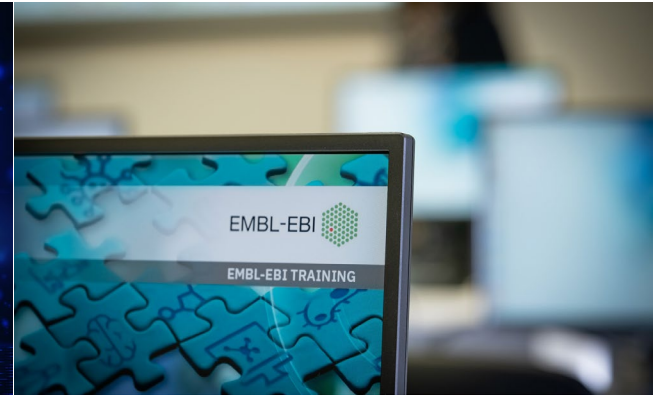
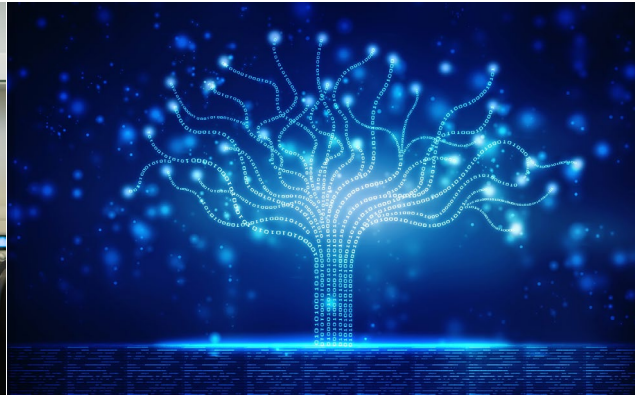
100 million
requests to our websites
on an average day

2 million scientists
access our websites
every month

7 out of 10 users say
“not having access to EMBL-
EBI data resources would have
a major impact
on my work”



Want to learn more and develop your skills?



[EMBL's TeachingBASE](#)

[Tree of Life project tutorial](#)

[EMBL-EBI tutorials -
Bioinformatics for the terrified](#)

Thank you

<https://www.ebi.ac.uk/>

