

Highlights 2024

ebi.ac.uk

EMBL's European Bioinformatics Institute (EMBL-EBI)

Contents

Foreword	2
Who we are	3
2024 in numbers	4
Data resources	6
Research	14
Training and industry partnerships	18
Community and infrastructure	21
New leadership	28
Facts and figures	30
Governance, management and funders	34

© 2025 European Molecular Biology Laboratory

This publication was produced by the Communications team at EMBL's European Bioinformatics Institute (EMBL-EBI).

Cover illustration: Karen Arnott/EMBL

For more information about EMBL-EBI please contact:
comms@ebi.ac.uk

Foreword

EMBL-EBI is laying the foundations for the future of life sciences by making biomolecular research data FAIR (Findable, Accessible, Interoperable and Reproducible). These principles are the bedrock of a global ecosystem that empowers scientists to openly share their discoveries and harness the vast potential of interconnected data.

In 2024, EMBL-EBI sharpened its focus on enhancing the usability and reach of its open data resources, bluesky research and training. By actively engaging with diverse voices across disciplines and regions, we are helping to build global bioinformatics capacity in sustainable, meaningful and inclusive ways. We perform cutting-edge computational research in the life sciences, driving forward new methodologies and discoveries. Through collaborative efforts, accessible training, and cutting-edge innovation, we are supporting researchers everywhere to use data-driven biology and artificial intelligence to deliver transformative solutions in healthcare, agriculture, biodiversity and beyond.

This vital work is only possible thanks to the steadfast support of our funders, including EMBL's member states, UK Research and Innovation, Wellcome, the European Commission, and the National Institutes of Health. Together with them and our global collaborators, we are keeping the world's biomolecular data open and accessible for all.



Ewan Birney
Director of EMBL-EBI and
EMBL's Deputy Director General



Jo McEntyre
Deputy Director
of EMBL-EBI

Who we are

EMBL's European Bioinformatics Institute (EMBL-EBI) is the world's leading source of public biomolecular data. We enable life science research and its translation to medicine, agriculture, industry and society by providing biological data, tools and knowledge.

We are one of the six sites of the European Molecular Biology Laboratory (EMBL), Europe's leading life sciences organisation. EMBL conducts research, provides training and state-of-the-art research infrastructures. EMBL is an intergovernmental organisation with 29 member states, one associate member state, and two prospect member states. EMBL's scientific programme – Molecules to Ecosystems – seeks to better understand life in context.

Our vision

To benefit humankind by advancing scientific discovery and impact through bioinformatics.

Our missions

- To freely provide data and bioinformatics services to the scientific community in ways that promote scientific progress.
- To contribute to the advancement of biology through investigator-driven research in bioinformatics.
- To provide bioinformatics training to scientists at all levels.
- To disseminate cutting-edge technologies to industry.
- To support, as an ELIXIR Node, the coordination of biomolecular data provision in Europe.



2024 in numbers

HR

657 

established members
of personnel*



from 68
countries

*in Full Time Equivalent includes staff, fellows,
supernumeraries and trainees during 2024 calendar year

DATA RESOURCES

123 million requests to our data
resources on an average day



from 42 million unique IP addresses annually



>110,000 scientific publications enabled
by EMBL-EBI data resources

RESEARCH



241
journal papers
published

INDUSTRY

1,696


attendees

to our 11
Industry Programme
workshops

GRANTS

185



active grants

154



collaborative grants
with researchers



in 64
countries



from 955
institutes

TRAINING



468,607
users accessed our
online training



107,429
webinar views



21
live courses



reaching 642
delegates

PUBLIC ENGAGEMENT

897



participants

23



public engagement
events

supported by 48



staff and
students

7



events organised with
our community partners

Data resources

EMBL-EBI manages the world's most comprehensive suite of open data resources for the life sciences. Our 40 data resources and dozens of tools span genetics, genomics, proteins, bioimaging, chemistry, literature and more.

Fundamental data infrastructure

The life sciences depend on high-quality data from public data resources, such as the ones managed by EMBL-EBI and collaborators. As the volume, complexity and usage of biodata continue to expand, EMBL-EBI is committed to engaging its user communities to understand their evolving needs, and to adapt its resources and tools, facilitating innovative applications of large-scale biodata.

In 2024 EMBL-EBI ran a user survey, which had over 2,300 responses from 126 countries. The survey revealed:

- 89% of respondents said that EMBL-EBI data resources empowered them to undertake work that would otherwise not have been possible.
- 73% of respondents reported it would not be practical to recreate or regenerate the data they access through EMBL-EBI.
- An increase in the number of respondents working in the government, hospital and not-for-profit sectors.



"The data resources we co-manage with our collaborators are used worldwide by an increasing number and breadth of people. Although user surveys have limitations, it's encouraging to see the many ways in which open data resources support research and discovery, from everyday computational analyses to training the next generation of AI models."

Jo McEntyre, Deputy Director of EMBL-EBI



Read more

Understanding genetic variants linked to disease

To improve patient outcomes and develop better treatments, researchers are working to understand which genetic variants are associated with diseases. Despite recent technological advancements, collating and assessing all of the relevant evidence remains a challenge. EMBL-EBI has taken steps to support these efforts:

- Integration of the AlphaMissense tool, developed by Google DeepMind, into the Ensembl, UniProt, ProtVar and AlphaFold data resources. AlphaMissense uses protein structure and sequence data to estimate how likely it is for a genetic variant to be pathogenic. Although AlphaMissense does not predict the impact of a variant on specific diseases, it can help researchers understand which regions of a protein are more likely to be pathogenic.
- The Ensembl Variant Effect Predictor (VEP) has integrated new pathogenicity predictions and data from external sources, including Open Targets. Widespread use of Ensembl VEP in national genomic medicine services means these updates could help improve patient care.
- The DECIPHER database for rare diseases launched an interactive tool to guide users through genetic variant classification. This enables easy access to evidence supporting or refuting variant pathogenicity, which speeds up analysis and improves consistency in how these criteria are applied in the clinic.
- The UK Association for Clinical Genomic Science now recommends the use of Gene2Phenotype (G2P) data in assessing potentially clinically relevant variants. G2P is a publicly-accessible online system co-developed by EMBL-EBI in collaboration with consultant clinical geneticists. It facilitates the validation, curation and distribution of large-scale, evidence-based datasets for use in genetic analysis.



"EMBL-EBI data resources work closely with their user communities to add new features that empower scientists to gain new insights from data. We're pleased to expand our offering for exploring genetic variation."

Maria Martin, Team Leader at EMBL-EBI

Secure access to sensitive human data across borders

In the era of genomics, data reuse is more important than ever, but sensitive human data should only be accessible through safe and secure environments. The Federated European Genome-phenome Archive (FEGA), coordinated by EMBL-EBI and the Centre for Genomic Regulation in Barcelona, is the primary global resource for discovery and access of sensitive human omics and associated data consented for reuse.

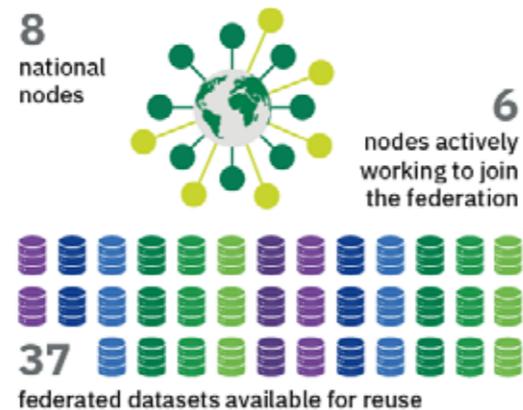
FEGA aims to accelerate disease research and improve human health. In 2024, FEGA operated through a network of eight national nodes (Canada, Finland, Germany, Norway, Poland, Portugal, Spain and Sweden), and marked a major milestone as the first datasets became available through the network.



“The value [of FEGA] for users is twofold; it becomes easier to make data available securely compared to storing data locally, and data becomes easier to find for other researchers. The possibility of reuse improves the chances of verifying scientific discoveries and reaching groundbreaking conclusions that benefit the research community and patients.”

Anna Hagwall, Head of Human Data at FEGA Sweden, National Bioinformatics Infrastructure Sweden

FEDERATED EGA IN 2024



Seeding a path for more efficient drug discovery

Identifying drug targets is a complex yet crucial process in the life sciences. Open Targets, an innovative partnership between academia, including EMBL-EBI, and the private sector, celebrated ten years of collaboration in 2024.

Open Targets aims to systematically identify and prioritise drug targets with a focus on immunology, cancer, and neurodegeneration, with the ultimate goal of helping drug discovery scientists develop safer and more effective treatments for disease.

In the decade since its inception, Open Targets has conceived and funded over 90 projects that have nominated novel promising targets, such as the DNA helicase WRN for specific colorectal cancers. The partnership has also created tissue atlases foundational for drug discovery research, launched powerful databases such as the eQTL Catalogue, and established Open Targets Genetics and the Open Targets Platform.

With over 70,000 public users in 170 countries, and over 20 customised instances, the Open Targets Platform has become a leading tool for the prioritisation of targets.



“The key to our success has been strong and ongoing collaboration between world-class academic institutes and leading pharmaceutical companies, and we are excited to build on this with even more ambitious projects in the decade to come.”

Ellie McDonagh, Translational Informatics Director of Open Targets



10 YEARS OF OPEN TARGETS



70,000 public users



170 countries



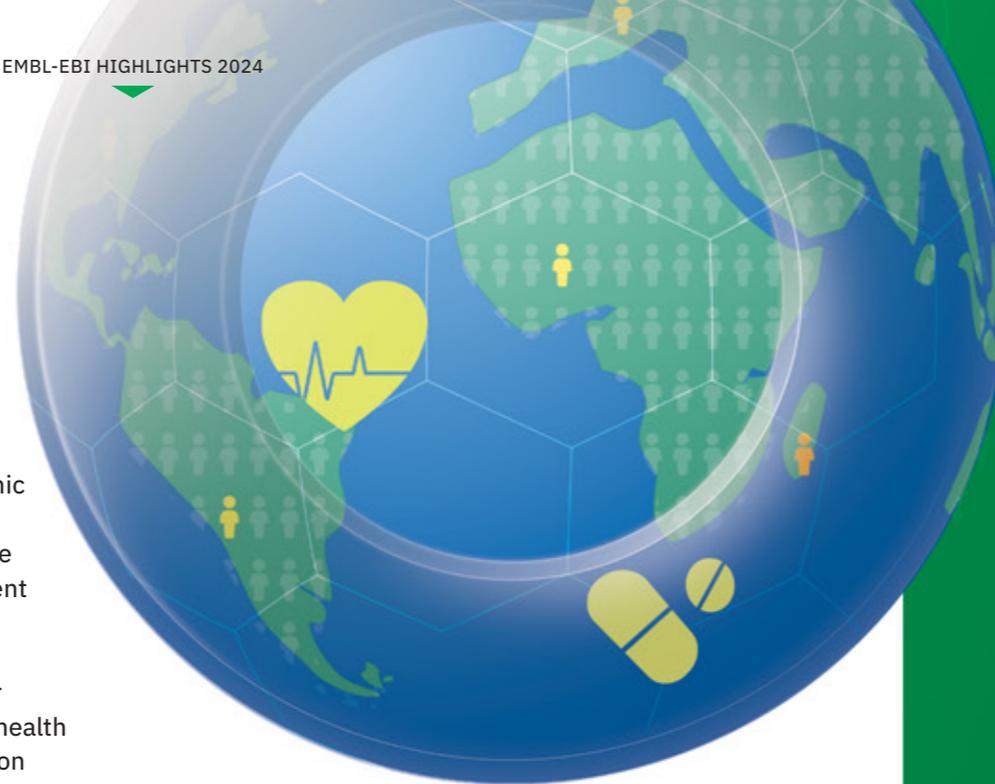
90 projects nominating promising drug targets

Advancing equitable disease predictions

The PGS Catalog is the largest open database for polygenic scores, with over 27,000 users from 140 countries. Polygenic scores estimate an individual's genetic predisposition to a specific trait or disease by summarising the effect of many different genetic variants across the genome.

Polygenic scores are particularly useful for predicting the risk of developing complex health conditions such as heart disease or common cancers, where multiple genetic variants contribute to the overall risk.

The PGS Catalog, co-developed by EMBL-EBI and the University of Cambridge, is standardising the way polygenic scores are reported, and making them more reliable for clinical applications.



The resource now features more data from non-European populations, as well as a new PGS Catalog Calculator. These improvements make it easier to calculate scores in different computing environments and across ancestries.

AYBUKE KUPCU YOLDAS

BIOIMAGE ARCHIVE DATA PROJECT LEADER

Originally from Turkey, Aybuke swapped telescopes for microscopes and stars for cells a few years ago, when she pivoted from astrophysics to bioimaging.

Joining EMBL-EBI in 2022 as an imaging bioinformatician, Aybuke was one of the first team members working on the BioImage Archive, EMBL-EBI's public repository for biological images from all modalities and scales.

Aybuke and her colleagues are helping to make bioimaging data easily accessible and interoperable, so researchers can gain new insights. They are also linking bioimages with other rich omics datasets to give scientists more biological context.

"Openly sharing biological images is still a relatively new practice, so we're working closely with our user communities to support their needs and to build a sustainable and scalable data sharing ecosystem," said Aybuke.



CASE STUDY

Proteomics: a different lens for precision medicine

Proteomics is shaping the future of precision medicine and EMBL-EBI is helping to overcome some of the barriers standing in the way.

The proteome consists of all the proteins that exist in an organism at any given moment. Unlike the genome, which is relatively fixed throughout an individual's life, the proteome is constantly changing.

What can proteomics tell us?

Proteomics experiments measure levels of proteins in different physiological conditions and draw out useful biomarkers for illnesses. For example, protein-protein signalling deregulation can lead to cancer, while protein aggregation can play a role in neurological diseases. Proteomics provides a useful lens through which researchers can see disease progression at a molecular level. It also uncovers useful insights for developing new clinical interventions.

Access to experimental facilities

Proteomics uses a wide range of highly complex technologies, such as mass spectrometry. Many of these technologies are available to scientists worldwide through EMBL's Proteomics Core Facility in Heidelberg, Germany, which supports over 400 scientific projects every year.

"In my view, EMBL-EBI's PRIDE data resource has been a real game-changer for proteomics and has brought the field to maturity."

Mikhail Savitski, Head of the EMBL Proteomics Core Facility

Improved data access

As proteomics technology has developed, the volumes of proteomics data have increased. Proteomics data sharing has become the norm thanks to:

- an increased adoption of open science practices
- the development of reliable data repositories, like EMBL-EBI's PRIDE, which enable data access and reuse
- requirements from journals and funders for researchers to share their data.

During 2024, the ProteomeXchange consortium, led by EMBL-EBI's PRIDE, reached an impressive 50,000 datasets, and the growth shows no sign of slowing.

Proteomics continues to change at a rapid pace. The public availability of proteomics data makes it a promising field for training new AI methods.

Read more



Ancient DNA provides solutions for global food security

Climate change is creating global challenges for agriculture, including rising temperatures and increased frequency of extreme weather events such as droughts and floods.

The Novo Nordisk Foundation and Wellcome have awarded funding for a new research programme called the Ancient Environmental Genomics Initiative for Sustainability (AEGIS), which includes EMBL-EBI. AEGIS will analyse ancient environmental DNA to gain insights into ancient crop genetic diversity and past adaptations to climate change, in the hope that these learnings can inform the development of more resilient crops.

EMBL-EBI plays a multifaceted role in AEGIS, acting as a data distributor, as well as supporting the data processing and analysis. EMBL-EBI will also develop a dedicated AEGIS public data portal, and add project data in existing resources:

- The European Nucleotide Archive will host raw genomic data and the team will develop metadata standards for working with ancient environmental DNA.
- The Ensembl team will annotate reference genomes to help identify what species the samples correspond to.
- The MGnify team will support the project by constructing ancient metagenomes from the samples collected.



"Agricultural systems and the foods we grow have changed over millennia. [...] AEGIS hopes to build the evidence base needed to strengthen the resilience of food systems across the world in the face of future environmental changes."

John-Arne Røttingen, CEO, Wellcome



CASE STUDY

Discovering new enzymes to upcycle food waste

The United Nations predicts that the global population will peak before the end of the century at just over 10.4 billion people. To feed this growing population, the world needs new approaches for sustainable food production.

A currently underutilised source of nutrients is animal by-products, including bones and skin. For instance, about 45% of a salmon's total weight currently goes to waste. Researchers are using the vast protein space in EMBL-EBI's microbiome data resource, MGnify, to identify new enzymes that can help extract animal proteins from by-products, in a robust and sustainable manner.

Scientists at NORCE, an independent research institute in Norway, found a solution at the bottom of the ocean. They found that some microbes that live on the seafloor can 'eat' the bones of dead whales, extracting nutrients, thanks to certain enzymes with bone-degrading properties.

This is an ingenious enzyme ability, but it does come with some limitations, including sensitivity to heat. The researchers used EMBL-EBI's MGnify database to try to identify a suitable cocktail of enzymes for extracting nutrients from animal by-products.

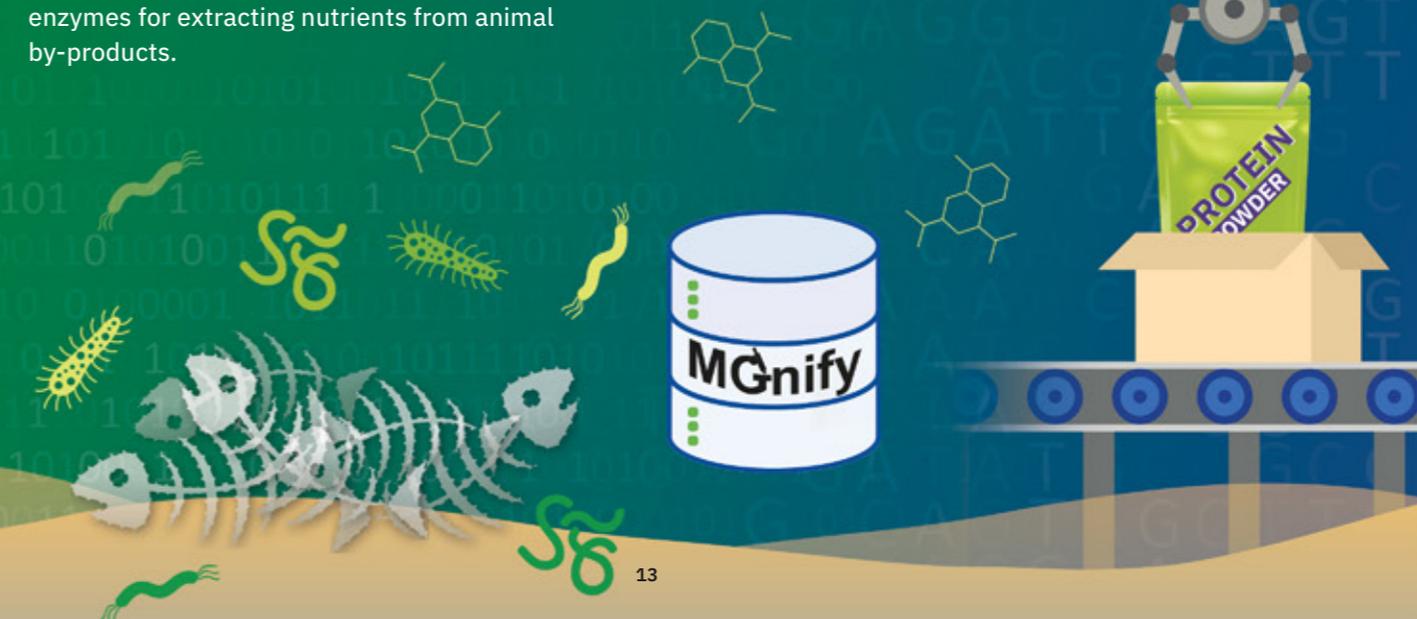
This work is part of the BlueRemediomics project, co-led by Rob Finn at EMBL-EBI. BlueRemediomics aims to systematically catalogue marine microbe data to facilitate the development of industrial processes that reduce waste, increase the reuse of natural and man-made products and by-products, and improve aquaculture processes.

Bone degradation is just one of many exciting applications explored by BlueRemediomics, which brings together experts from different disciplines. With the help of Christine Orengo's team from University College London (UCL), the team is leveraging artificial intelligence to search huge volumes of data to find the most suitable enzyme candidates.

"MGnify will help us to identify the right enzyme or cocktail of enzymes with the traits required for a specific bioprocess."

**Antonio Garcia-Moyano,
Senior Researcher at NORCE**

Read more



Research

EMBL-EBI's research groups use computational methods and artificial intelligence to make sense of vast, complex datasets. Our researchers collaborate with experimental scientists to build practical solutions for real-world challenges linked to human health and the environment.

Predicting cancer risk on the basis of national health data

Detecting cancer early gives patients more treatment options and generally results in better clinical outcomes. But current screening programmes have limitations; they focus on specific cancer types and don't always include the individuals most at risk.

Researchers in the Birney and Gerstung groups at EMBL-EBI and DKFZ used data from Danish health registries to create a statistical model that can predict an individual's risk of developing 20 different types of cancer, based on their family history, disease history, and lifestyle. The model doesn't produce an exact prediction of which person will develop cancer, but it does determine the individual risk, and enables a comparison with people of a similar age.

"Such models rely on a system for capturing and leveraging population-wide health data. The Danish health data ecosystem holds digital data for the entire population and spans decades. Only a few European countries offer something similar, including Finland, Iceland and Sweden, or special research cohorts in the UK."

Moritz Gerstung, Division Head at DKFZ and EMBL-EBI Alumnus

The team also tested the model using UK Biobank data and achieved comparable levels of accuracy, which suggests the method could be applied across national health systems.

The work confirmed well-known factors associated with cancer, such as smoking and alcohol consumption; it also found that while family history is most informative before the age of 45, as individuals age, their disease history becomes more relevant to their cancer risk.

Addressing cancer disparities in African heritage populations

Prostate, breast, and pancreatic cancers disproportionately affect people of recent African heritage, manifesting as more aggressive tumour types and with earlier onset compared to other populations. This disparity highlights a critical health issue and an urgent need for targeted research and intervention strategies.

The Cancer Grand Challenges initiative has awarded funding to an international team of researchers, to decode the factors underpinning cancer inequities among recent African heritage populations. Team SAMBAI includes researchers from Ghana, South Africa, UK, USA and EMBL-EBI.

The researchers will integrate different data types, including genomics, exposomics, immune profiling, and social determinants of health data. The aim is to identify factors leading to disproportionate cancer rates and worse outcomes in people of recent African descent. The project will also support the development of molecular profiling and training capacity in Africa, and establish new methods to study cancer inequities in other underserved populations.

Cancer Grand Challenges is an initiative designed to unite the global research community to tackle the most complex and persistent cancer-related problems through collaborative approaches. The funding for team SAMBAI comes from Cancer Research UK and the National Cancer Institute in the USA.



"By collaborating with local scientists and institutions, we aim to better equip communities with the tools and knowledge necessary for ongoing cancer research."

Isidro Cortes-Ciriano, Research Group Leader at EMBL-EBI

Understanding the genetic flexibility of the malaria parasite

Plasmodium falciparum is the deadliest malaria parasite in humans. By understanding its genetics, researchers can deepen their knowledge of how the parasite evolved, and could inform vaccine development.

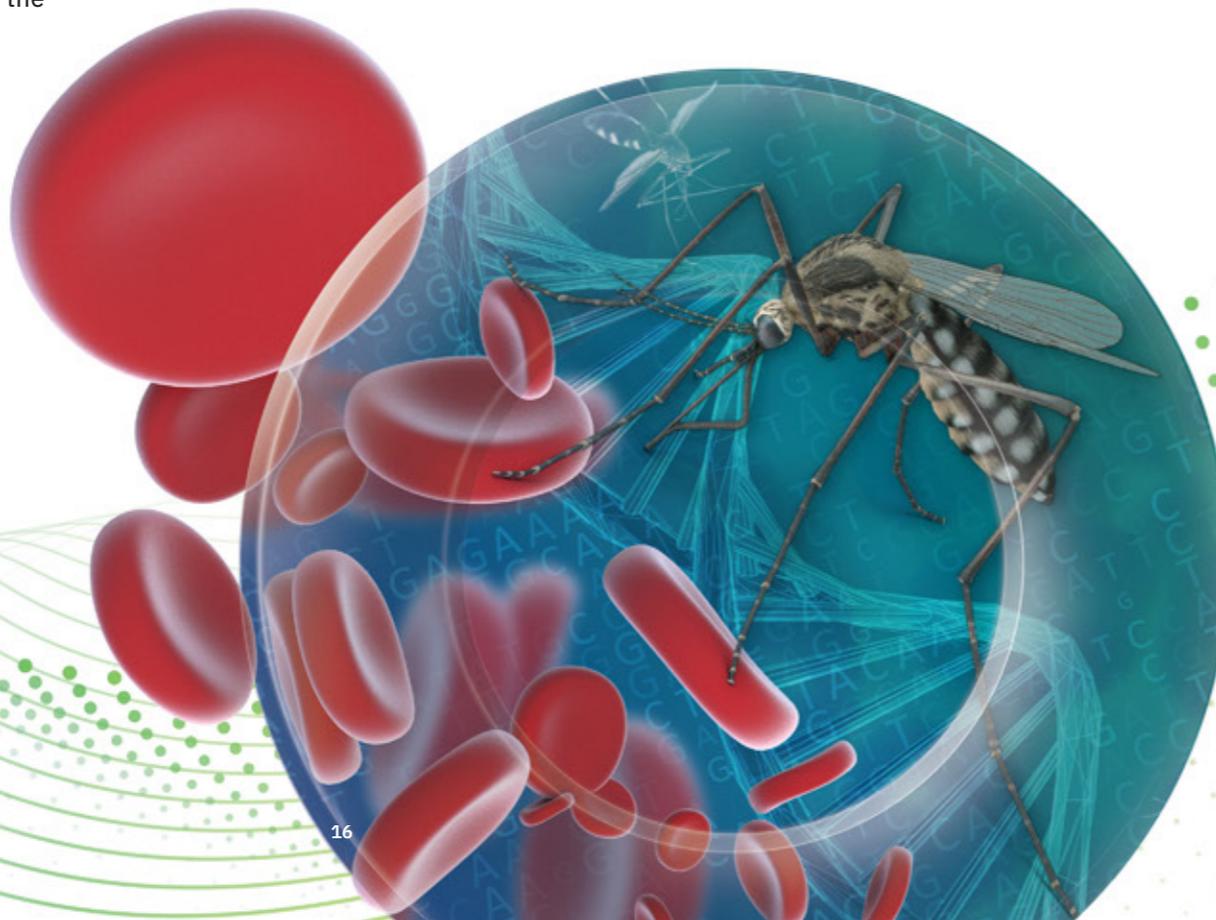
Researchers at EMBL-EBI have identified two unusual hotspots of genetic diversity in genes DBLMSP and DBLMSP2. These seem to be the result of a 'copy-paste' mechanism that increases the genetic diversity of the parasite at accelerated time scales. This insight helps solve the long-standing mystery of why the parasite displays hotspots of genetic diversity in an otherwise unremarkable genetic landscape.

The two hotspots are so extreme that current algorithms for mapping genetic variants failed to capture them. By developing new bioinformatics software that uses genome graphs and analysing data from 29 countries, the researchers managed to reveal previously hidden variants, and demonstrate that multiple gene conversion events had occurred.



"We have been trying to understand the unusual patterns in these genes for almost a decade. These data not only enhance our grasp of malaria's biology, but will also be valuable to researchers across the world studying these genes and their interaction with our immune system."

Zamin Iqbal, Visiting Group Leader at EMBL-EBI and Professor at the University of Bath



FRANCESCO COSTA PREDOCTORAL FELLOW



After studying biotechnology in Italy and Germany, Francesco Costa joined the EMBL international PhD programme to study proteins.

He is fascinated by protein design, which is the process of engineering new proteins or modifying existing ones to perform specific functions. Protein design has applications in medicine, energy and materials, among other sectors.

Working with EMBL-EBI services teams, Francesco has developed a method that helps to improve the confidence score for many AlphaFold 2 protein structure predictions. He is also researching fibrillar adhesins – proteins used by bacteria to attach to surfaces and interact with other bacteria.

"EMBL has an incredible ability to bring people together and create opportunities to exchange knowledge and ideas. This is extremely valuable, especially for early-career researchers," said Francesco.

MARIA HERRERO-ZAZO POSTDOCTORAL FELLOW



While working in a community pharmacy in Spain, Maria decided to pursue a career in academia. She obtained a PhD in computer science in Madrid, followed by a postdoc at King's College London. She has now completed a second postdoc at EMBL-EBI.

Maria's EMBL-EBI work was a proof of concept that machine learning models could be used to gain insights from clinical records, to help understand why some patients respond to drugs while others don't.

At EMBL-EBI, Maria deepened her genetics expertise by analysing data from UK and Danish health registries and biobanks to explore patient response to antidepressants.

"I appreciate the flexibility to design my own projects and explore research leads. My work always comes back to patients and how we can use the data we have to gain useful insights into prescriptions, clinical trajectories and drug response," said Maria.

Training and industry partnerships

EMBL-EBI delivers world-leading training in biodata science for all levels, including predoctoral and postdoctoral. Our approach to democratising global life science research and building bioinformatics capacity is reflected in our training programmes.

Democratising Nobel Prize winning technology

Proteins are essential components of all living things and predicting their 3D structures offers valuable insights into how life works. In 2024, the Nobel Prize for Chemistry was jointly won by Google DeepMind's Demis Hassabis and John Jumper – the creators of the AlphaFold AI system, which predicts the 3D structures of proteins – and David Baker at the University of Washington, Seattle, for his work in computational protein design.

EMBL-EBI's long-standing collaboration with Google DeepMind made the AlphaFold 2 protein structure predictions freely-available to the global scientific community through the development of the AlphaFold Database in 2021.

In 2024, in response to our user community needs, EMBL-EBI and Google DeepMind co-created a free, open-access, comprehensive online course for AlphaFold 2. The course enables life scientists, including those with limited experience in structural biology, to incorporate structure prediction using AlphaFold into their research projects responsibly and efficiently. The feedback from learners has been extremely positive, and

EMBL-EBI encourages re-use and adaptation of the course materials, to fit appropriately into the training and education programmes of others.

In the first 9 months since launch, the AlphaFold course had

- Over 630,000 page views
- Almost 67,000 users from 159 countries

"The AlphaFold training course helps me understand how AlphaFold works and how to use the output in a responsible manner."

Trainee feedback

The AI revolution hinges on accessible training

Sameer Velankar oversees the Protein Data Bank in Europe team at EMBL-EBI, which manages the AlphaFold Database. Here, he explains how training helps users make the most of AI technologies.



Why is training important in the life sciences?

Accessible training lowers the barriers to entry and enables life scientists around the world to integrate new technologies into their work effectively and responsibly.

Understanding how to use results from new technologies or databases is not straightforward; a healthy amount of background knowledge and critical thinking are usually required to assess whether the data you get from new technologies are correct and useful in a given context.

It's also important for users to be aware of the limitations of technology – what it can and can't do, what it's good at, and where it falls short. This is only possible through robust documentation and accessible training.

How would you describe training that is accessible?

Accessibility is multifaceted. At a minimum, training should be easy to find and not behind a paywall. EMBL-EBI's Training team has a long history of providing freely available training in an electronic format so it can be accessed by a global audience free of charge.

Accessible training is comprehensive and easy to understand by different users with a variety of training backgrounds, levels of expertise and abilities. The only way to navigate this is to continually engage with the community, taking into account feedback and questions.

Why do AlphaFold users need training materials now?

Until a few years ago, the availability of protein structure data was limited so not everyone needed to know how to use structure data effectively. But since Google DeepMind and EMBL-EBI made millions of AlphaFold protein structure predictions public, we have entered a world where structural data is abundant.

Now, everyone can access relevant protein structure models, whether they are studying human health, crops, biodiversity, enzymes, or something else entirely. AI predictions won't replace experimental data, but they are a useful tool that the scientific community has been using heavily and creatively. There are already 18,000 scientific papers citing AlphaFold, and the database has over 2 million users in 190 countries.

AlphaFold predictions are reaching millions of users who have never had much contact with protein structure data. 'AlphaFold: a practical guide' aims to fill the gap and support scientists worldwide wanting to use this rich dataset.

KIM GURWITZ**SCIENTIFIC TRAINING LEAD**

Kim started her career as a medical biochemistry researcher, but quickly realised that what she enjoyed most was facilitating learning experiences for others. As a training coordinator, she supported learning and development across Africa and Europe through initiatives such as H3ABioNet, a pan-African bioinformatics network.

At EMBL-EBI, Kim works in partnership with user communities to identify training gaps, source funding, and develop training programmes, and leads a team to deliver on this work. A notable project is the Pathogen Data Network, a global consortium enabling FAIR infectious-diseases data sharing and reuse.

“EMBL-EBI’s approach to training is based on mutual respect and co-design. We work closely with subject matter experts and scientists in our user communities to co-develop training that builds competence, capacity, and culture. This way, we learn from each other and build long-standing relationships,” said Kim.

Expanding our collaboration with industry

EMBL-EBI has a mission to disseminate cutting-edge technologies to industry. We support companies of all sizes that want to harness the power of bioinformatics for research and development.

Our recently launched Agri-Tech Partnership involves working with industry and academic partners to optimise the utility and value of publicly available biological data. The aim is to build a fit-for-purpose AgriData Platform integrating EMBL-EBI data and other public data sources worldwide.

In 2024, the Agri-Tech Partnership welcomed its first signatories: Unilever, Bayer, Syngenta and Biotalys.

We welcome new partners who can:

- Share their use cases and determine which datasets to prioritise
- Make us aware of gaps in existing public data and tools
- Collaborate on projects to close gaps or to generate new data relevant for their needs

Community and infrastructure

To fulfil its missions, EMBL-EBI relies on a deeply collaborative culture, a diverse workforce and robust technical infrastructure.

Thornton Building: a new hub for collaboration

In November 2024, EMBL-EBI moved into its new Thornton Building, which was made possible by generous support from the United Kingdom Government via UK Research and Innovation (UKRI), as well as Wellcome, and the Biotechnology and Biological Sciences Research Council (BBSRC).

Named after the institute’s Director Emeritus and esteemed scientist Janet Thornton, the building is a space for collaboration and translational research. The new space will help to turn EMBL-EBI’s data management expertise into practical solutions for global challenges such as infectious disease, food insecurity and loss of biodiversity.



Janet Thornton - a bioinformatics pioneer

Professor Dame Janet Thornton is a structural bioinformatics pioneer. Her research focuses on characterising and analysing protein structures, and understanding how proteins contribute to disease and ageing. During her tenure as EMBL-EBI Director between 2001 and 2015, the institute grew from 160 to 600 people, in no small part thanks to Janet's drive and leadership style.

Janet Thornton was appointed Dame Commander of the Order of the British Empire in 2012 for services to bioinformatics. During her career she has held governance roles in the Royal Society, European Research Council and ELIXIR. Janet Thornton retired in 2023 but continues to be a regular visitor at EMBL-EBI.



This building is for all those people who, over time, have worked to make EMBL-EBI such a special place. I hope this building will enable everyone at EMBL-EBI to continue doing what they do best, which is to deliver data and help to interpret it. This is a very special remit and we are unique in the world because of it."

Dame Janet Thornton

Inspired by nature, designed for collaboration

The Thornton Building was designed with collaboration and environmental sustainability in mind. This approach aims to break down silos and bring together research, services and technology to encourage cross-team working. There are many shared spaces and facilities for collaboration, including a new seminar suite with breakout rooms and an auditorium-style space.

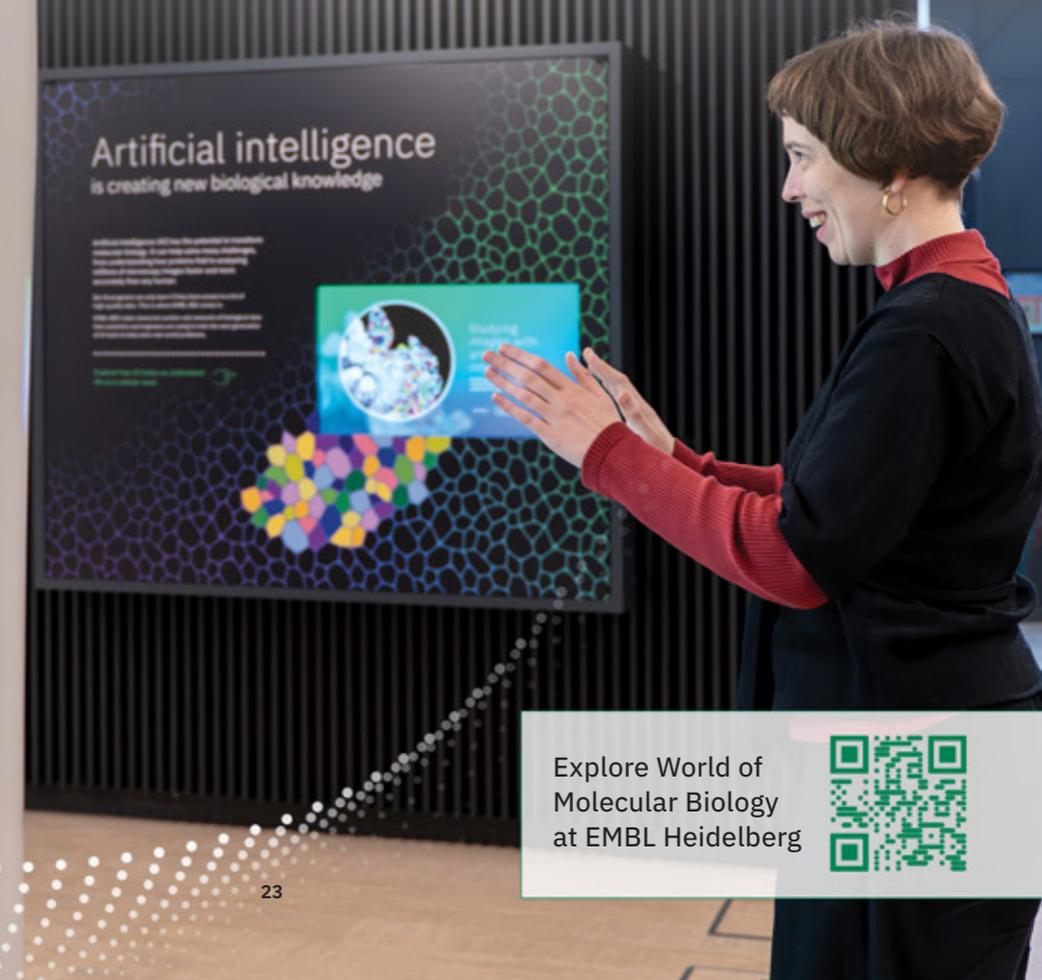
The Thornton Building design boasts biophilic principles including an abundance of natural light, the use of sustainable materials, such as wood and cork, and indoor planting. Incorporating natural elements into the building design, creates a healthier and more appealing working environment.



Dedicated wellbeing and public engagement spaces

The building is equipped with a multifaith/wellbeing room, a new parents' room, showers and bicycle parking, as well as EMBL-EBI's first dedicated IT Support Desk.

There is also a public engagement space inspired by the World of Molecular Biology exhibition at EMBL Heidelberg, with interactive exhibits that introduce visitors to the fascinating worlds of bioinformatics, genomics and artificial intelligence.



Explore World of
Molecular Biology
at EMBL Heidelberg



A sustainable building



POWERED BY GREEN ENERGY

The Thornton Building uses renewable energy. Solar panels supplement the electricity supply and an air source heat pump provides heating throughout.



SUSTAINABLE CORK FLOORING

Our cork flooring is eco-friendly - no need to cut down the trees, just use the bark.



PROTECTING LOCAL WILDLIFE

We installed bat and bird boxes on campus, as well as bee bricks and bug hotels nearby.



TREE PROTECTION

We moved four hornbeam trees to the campus woodland area to make space for the building. They survived the move and are showing new growth.



TIMBER FRAME

The building has one of the largest timber frames in the UK (1000 m³). This helped to significantly reduce the building's carbon footprint.



ENERGY EFFICIENCY

Our heating and ventilation system reuses waste heat, boosting efficiency and saving energy.



REUSE OF DUG UP SOIL

All 27,000 m³ of soil dug up for foundations was reused on-site, for landscaping. No transport costs, low carbon footprint.



SUSTAINABILITY STANDARDS

Our BREEAM Excellent score of 76% shows a high standard of sustainability in energy, water, materials, and wellbeing.

ZOE ROBINSON

FACILITIES OFFICER



Recruited from the local area of the Wellcome Genome Campus, Zoe has a background in facilities management. Crucially, she has a knack for understanding the office needs of her colleagues, often before they do.

"The Facilities team ensures EMBL-EBI's buildings are in working order. If there are any issues, we aim to deal with them as soon as possible," said Zoe.

This covers everything from building management systems, such as water, lighting, heating and ventilation, to furniture and staff wellbeing. The team works closely with Health and Safety colleagues, to make sure that the working environment is safe, comfortable and accessible for all.

Zoe spearheaded the work to ensure the Thornton Building was kitted out and ready for its first occupants in record time, while also coordinating a large-scale office move which included over 400 colleagues.

"I love the variety of my job, and I feel very lucky to be part of such a supportive and hardworking team," said Zoe.

Building bridges through public engagement

EMBL-EBI takes a co-developed and relational approach to public engagement meaning we work with partners, which tend to be charities working with specific groups. They help us develop interesting and impactful public engagement activities to illustrate our science. This collaborative approach recognises the knowledge and expertise the public has.

One of the events we organised in 2024 was a bespoke science fair in Bury St Edmunds for Suffolk Family Carers (SFC) young carers (5-25 years old) and their families. This was an opportunity for EMBL-EBI to build connections with these families and for them to find out about bioinformatics and the life sciences. The event is a culmination of a variety of engagement activities developed with, and supported by, the SFC team and EMBL-EBI colleagues.

Read more



"EMBL-EBI has been able to adapt to the needs of our young carers. It's not just another group coming to them [...]; they've listened to what we feel our young people want and adapted to it. It's so valuable because it shows our young people that there's a community out there that cares for them."

Keiron Whall, Advisor at Suffolk Family Carers

Enabling knowledge exchange

In 2024 EMBL-EBI hosted a number of delegations, including the London Diplomatic Science Club, welcoming counsellors and science attachés from Belgium, Canada, Croatia, France, Germany, Hungary, Netherlands, Norway, Poland and South Korea.

Representatives from the National Institutes of Health, UK Research and Innovation councils, the French Embassy in London and the UK Biobank also visited EMBL-EBI as part of our commitment to knowledge exchange.

Digitalising internal processes

EMBL's administration and operations teams have spearheaded a phased transition to a new enterprise resource planning system, Workday.

Phase one focuses on the modernising and streamlining the human resources function, by digitalising manual processes, enhancing data accuracy and reporting, and aligning HR practices with the organisation's strategic objectives.

The transition began in late 2024 and is expected to take four years to complete. The aim is to move to a more efficient, streamlined and interconnected way of working, bringing the human resources, finance and planning functions onto a single, unified platform.



SHRAYANS AGRAWAL NETWORK ENGINEER



Shrayans is originally from Rajasthan in India, and started his career in IT in 2006. His role is all about balancing connectivity and network security to ensure that everyone at EMBL-EBI can access the technical infrastructure they need in a seamless way.

Shrayans led the network set-up for EMBL-EBI's new Thornton Building. It was a team effort that saw 40 network switches installed in the building in record time - all at arm's reach and invisible to the eye.

"EMBL-EBI is the most wonderful place I've worked at in terms of location. I'm not a fan of tall buildings so our sprawling campus and its natural beauty make me smile every day," said Shrayans.

Developing world-class IT services

EMBL-EBI's data resources and research heavily rely on the institute's technical infrastructure. Without them, we would not be able to manage the world's biological data. The institute's technical infrastructure includes almost 500 petabytes of raw storage, a vast cloud and virtualisation estate, and a high-performance computing environment made up of 21,000 cores.

In 2024, the first Scientific Advisory Committee for IT and Technical services took place. The review was positive, describing EMBL-EBI IT as "world-class".

A data archive for the life sciences

As one of the main keepers of the world's biomolecular data, EMBL-EBI is constantly growing and developing its archives. One of the institute's central data archives, the File Replication Archive (FIRE) reached a record 100 petabytes in 2024. For over 15 years, FIRE has been enabling thousands of scientists to share and access data openly to advance their work in key areas such as human health and disease, and agritech.

The archive ensures that critical data resources remain available, performant and secure at all times. It also supports data preservation and continuity, which are essential for the life science community.

"FIRE's ability to seamlessly scale and support the ever-growing needs of the scientific community underpins EMBL-EBI's mission to advance life sciences research."

Andy Cafferkey, Head of IT & Technical Services

FIRE provides a home for some of EMBL-EBI's largest and most popular data resources, including the European Nucleotide Archive (ENA) and the Bioimage Archive among others.

New leadership



Jo McEntyre

Associate Director for EMBL-EBI Services, Jo McEntyre, was promoted to Deputy Director of EMBL-EBI. Alongside her existing responsibilities in leading the services, training and IT areas of the institute, she deputises for Director Ewan Birney.



Mallory Freeberg

Mallory Freeberg was promoted to Human Genomics Team Leader, supporting the delivery of open data resources such as Ensembl and DECIPHER, which exist to enable a better understanding of human health and disease.



Julio Saez-Rodriguez

Julio Saez-Rodriguez was appointed EMBL-EBI Head of Research, responsible for setting the vision, strategy, and scientific direction of the cutting-edge research taking place at the institute.



David Ocaña

David Ocaña was promoted to Head of Applications, in charge of EMBL-EBI's vast database estate, web application platform, and public cloud activities.



Mary Barlow

Mary Barlow was promoted to Head of Major Initiatives. She supports the Director of EMBL-EBI with key institute-wide planning and delivery requirements. She continues to lead the Major Initiatives team, which manages strategic projects, such as the large Data Infrastructure programme funded by the UK Government.



Tim Porter

Tim Porter was appointed Head of IT Operations, managing the team that runs the foundational technical systems that the institute relies on.



Christina Ernst

Christina Ernst was appointed as Functional Genomics Team Leader. Her team delivers tools and services for the archiving, analysis, and visualisation of data obtained from functional genomics experiments, including the Expression Atlas data resource.



Dan Trigg

Dan Trigg was appointed as EMBL-EBI's Head of Campus Operations. He manages EMBL-EBI's relationship with the Wellcome Genome Campus, and leads the EMBL-EBI's Facilities and Health and Safety teams.

Facts and figures

Staff numbers

Personnel categories during 2024*



578
Staff members

40
Postdoctoral fellows

29
Predoctoral fellows

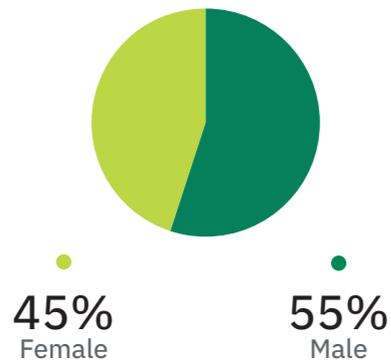
3
Supernumeraries

7
Trainees

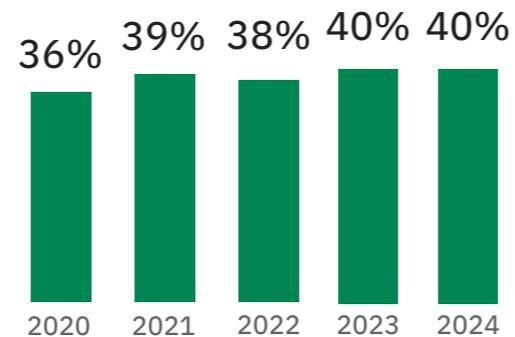
*Full-time equivalent (FTE)

EMBL-EBI hosted 160 visitors in 2024

Gender distribution of staff in 2024



Senior roles** held by women



**Grade 8 and above.

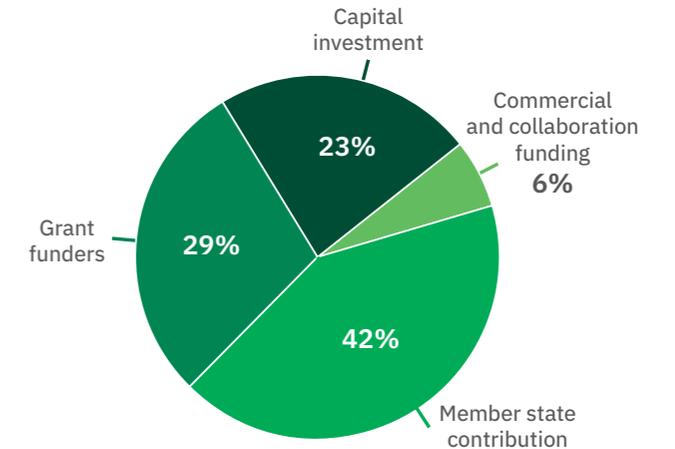
*These numbers cover EMBL-EBI and other EMBL staff in Hinxton during the entire 2024 calendar year.

Financial information

We are grateful to the EMBL member states and other funding bodies for their continued support. They enable us to maintain and grow our data resources, conduct vital research and training, and respond to the changing requirements of the global scientific community.

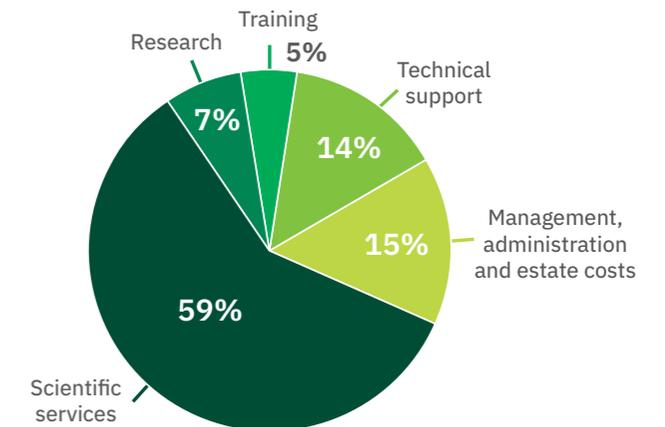
Funding sources

EMBL's financial year runs from 1 January to 31 December. EMBL-EBI receives its funding through four main funding channels: EMBL member state contributions (42%), commercial and collaboration funding (6%), external funding bodies via grant awards (29%) and capital awards for investment in buildings and data infrastructure (23%).



Operating expenditure

The total operating expenditure of EMBL-EBI in 2024 was €105.5m from all funding sources. Scientific services accounted for 59% of the expenditure in support of EMBL-EBI's data resources, with a further 7% allocated to research and 5% to external training. Approximately 14% of costs were on technical support which maintains and develops the institute's technical and IT infrastructure, and 15% on management, administration, and estate costs.



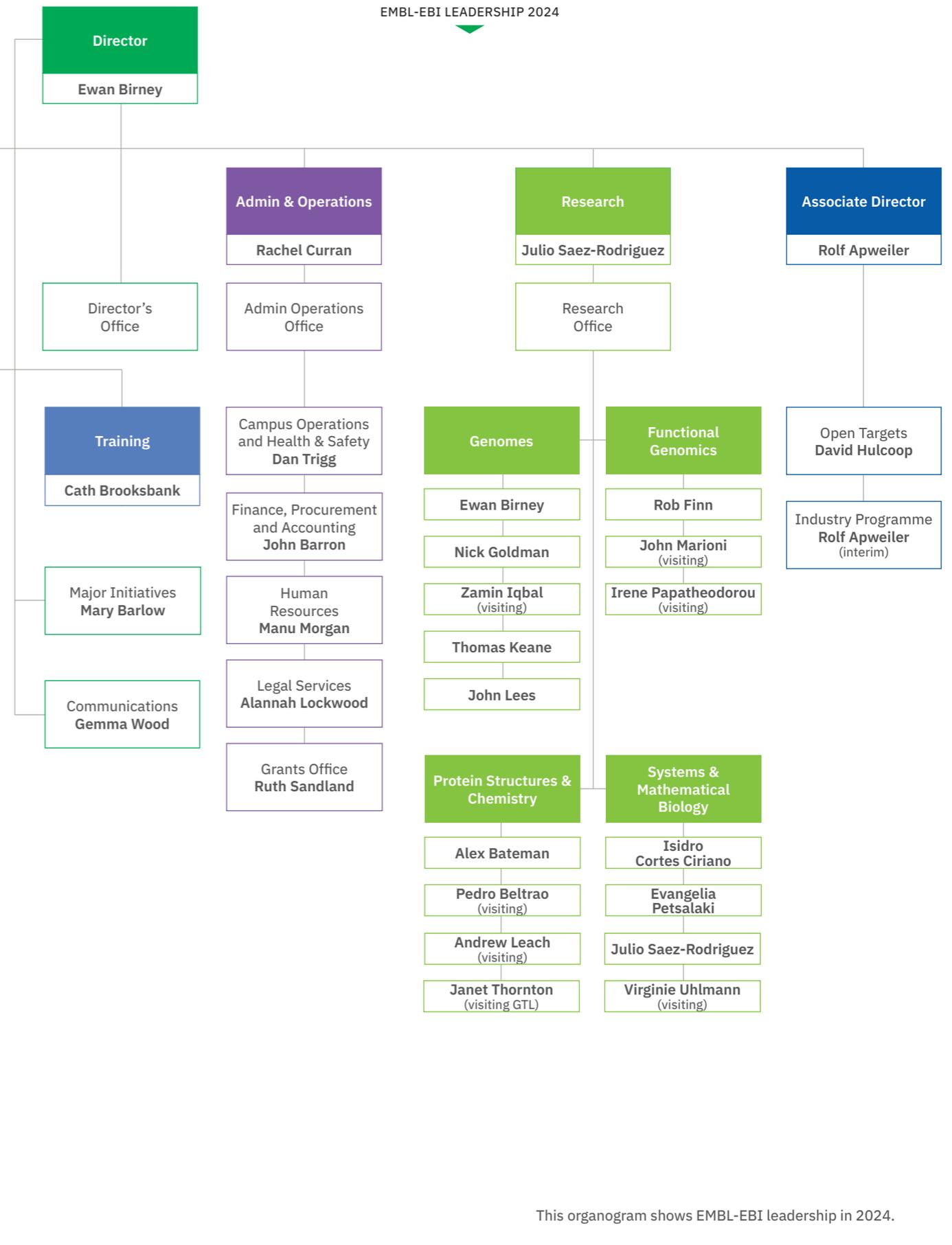
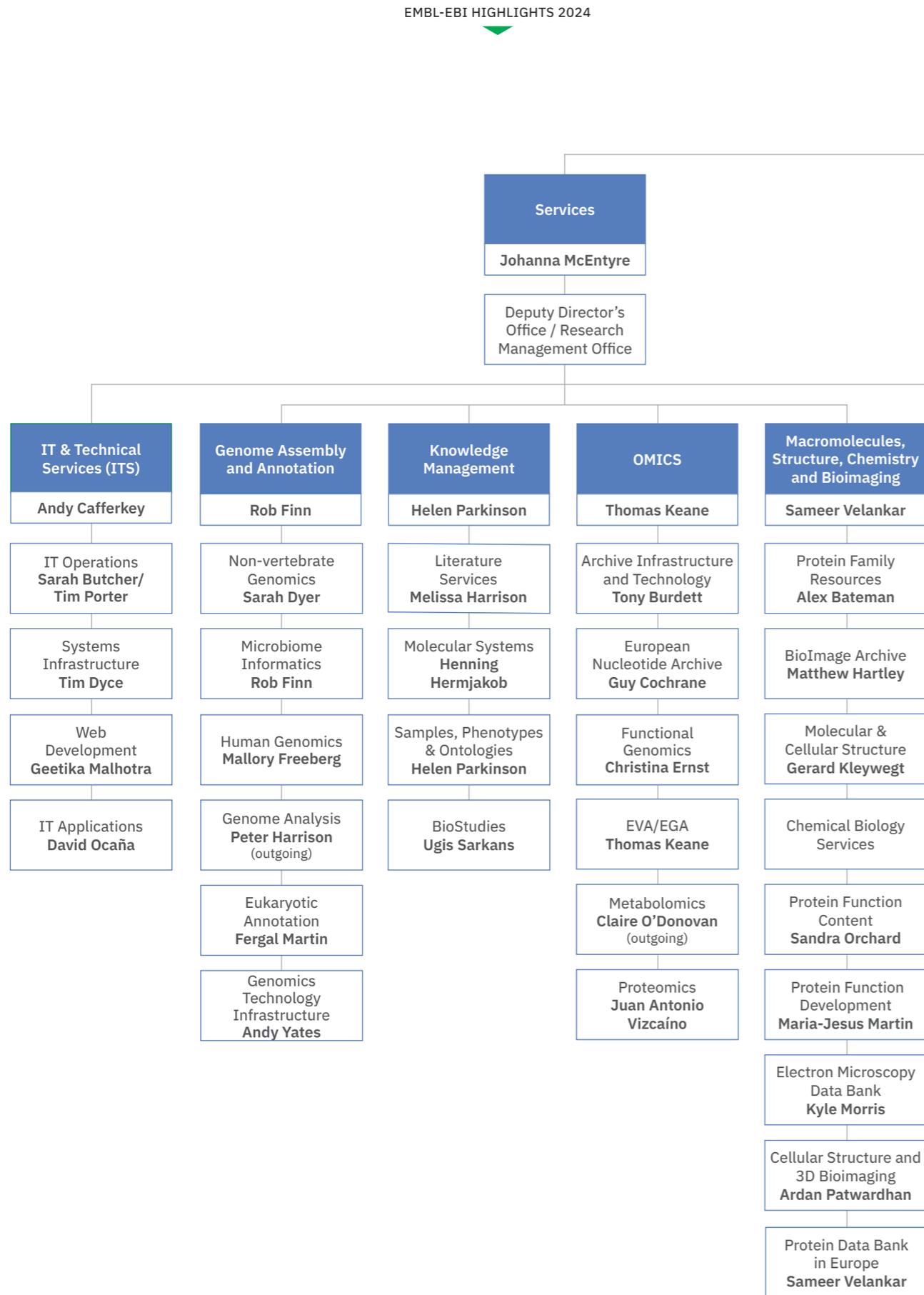
Capital investment

EMBL-EBI's Data Infrastructure Programme - Phase 2 (2024 - 2030) oversees £80.7 million of capital investment, which comes from the UK Government's UKRI Infrastructure Fund. The capital investment will support EMBL-EBI's continued technical infrastructure transformation (data storage, networking, and compute) in response to the growing need for open access biological data and expanded resource capabilities (data analysis tools, linkages, ontologies).

EMBL-EBI's Thornton Building, which opened in 2024, was funded by the UK Government via UK Research and Innovation (UKRI), as well as Wellcome and the Biotechnology and Biological Sciences Research Council (BBSRC).

Strategic partners and funders

EMBL-EBI works closely with strategic partners and funders from across the globe. Alongside funding from EMBL member states, in 2024 the institute received €39.9m in research and service grant funding from 24 funders providing support for 185 projects.



This organogram shows EMBL-EBI leadership in 2024.

Governance, management and funders

EMBL-EBI is part of the European Molecular Biology Laboratory (EMBL), an intergovernmental organisation with 29 member states, one associate member state and two prospect member states. EMBL is led by the Director General, Edith Heard, who is appointed by the EMBL Council.

The EMBL Council is composed of representatives from all member states of the Laboratory and determines its policy in scientific, technical and administrative matters by giving guidelines to the Director General. The Council ensures that the financial requirements of the agreement establishing EMBL, and of the agreements with host member states, are complied with.

In 2024, EMBL-EBI was led by the Senior Management Team composed of Director Ewan Birney, Deputy Director Johanna McEntyre, Head of Research Julio Saez-Rodriguez, Head of Administration and Operations Rachel Curran and Associate Director Rolf Apweiler.

We would like to thank the EMBL member states as well as our other funders listed below for their continued support.

- Biotechnology and Biological Sciences Research Council
- Bill & Melinda Gates Foundation
- Cancer Research UK
- Chan Zuckerberg Initiative
- European Association for Cancer Research
- European Commission
- Engineering and Physical Sciences Research Council
- Economic and Social Research Council
- Fonds National de la Recherche
- Leona M. and Harry S. Helmsley Charitable Trust
- Leverhulme Trust
- Mark Foundation for Cancer Research
- Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg
- Gordon and Betty Moore Foundation
- Medical Research Council
- Natural Environment Research Council
- NF1 Research Initiative
- National Institutes of Health
- National Institute for Health Research
- Novo Nordisk
- Osteosarcoma Institute
- Research Council RCUK
- Pezcoller Foundation
- Wellcome Trust

Imprint

Publisher
EMBL-EBI

Portrait image credits

p. 8 (Anna Hagwall): Anna Hagwall
p. 12 (John-Arne Røttingen): Wellcome

All portraits not listed here are from personal collections or taken on behalf of EMBL by Jeff Dowling.

Other image credits

p. 3 (Thornton Building exterior): Jeff Dowling/EMBL
p. 9 (Open Targets group): Phil Mynott
p. 18 (Gasdermin D AF-Q9D8T2-F1): AlphaFold DB/EMBL-EBI
p. 19 (Putative sucrose-specific porin AF-A0A809NSP9-F1): AlphaFold DB/EMBL-EBI
p. 21 (Thornton Building exterior): Jeff Dowling/EMBL
p. 22 (Thornton Building Forum): Jeff Dowling/EMBL
p. 23 (World of Molecular Biology exhibit): Jeff Dowling/EMBL
p. 24 (Thornton Building exterior): Jeff Dowling/EMBL
p. 25 (Public engagement event): Alex Brentnall/Suffolk Family Carers
p. 26 (Training activity): Jeff Dowling/EMBL

Printer

Healeys. healeys-printers.co.uk

EMBL's European Bioinformatics Institute (EMBL-EBI)

Wellcome Genome Campus
Hinxton, Cambridge, CB10 1SD
United Kingdom

 www.ebi.ac.uk
 +44 (0)1223 494 444
 comms@ebi.ac.uk

 ebi.embl.org
 @emblebi
 /EBImedia
 /company/ebi/

EMBL-EBI is a part of the European Molecular Biology Laboratory.

A digital version of this publication is available on
www.ebi.ac.uk/about/our-impact

EMBL member states and associate member states: Austria, Belgium, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom, Australia

Prospect member states: Bulgaria, Serbia