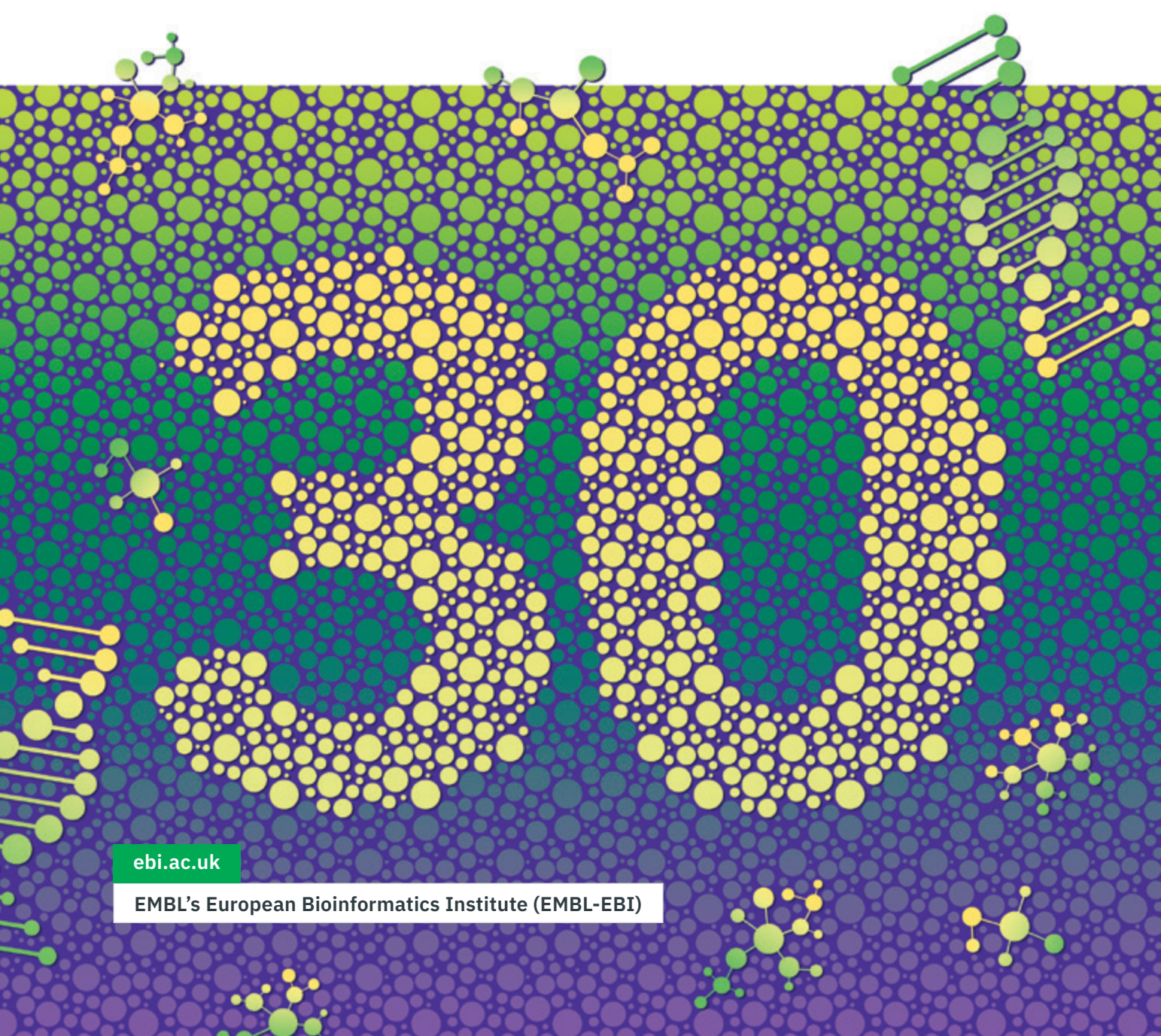


# Highlights 2023



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

EMBL's European Bioinformatics Institute (EMBL-EBI)



# Contents

Foreword	2
Who we are	3
2023 in numbers	4
Celebrating 30 years of excellence in bioinformatics	6
Data resources	8
Research	15
Training and industry partnerships	18
Culture and infrastructure	22
New leadership	27
Facts and figures	28
Governance, management and funders	29
EMBL-EBI leadership	30

© 2024 European Molecular Biology Laboratory

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

Cover illustration: Karen Arnott

For more information about EMBL-EBI please contact: [comms@ebi.ac.uk](mailto:comms@ebi.ac.uk)



# Foreword

When EMBL-EBI was set up in 1994, few could have predicted the central role that open data resources would play in the advancement of life sciences. Biologists have always championed research data sharing, but the level at which this is now happening around the world is truly staggering.

The advancements that open data and bioinformatics have enabled are astonishing, from the development of safe vaccines in record time, to the creation of revolutionary AI algorithms such as AlphaFold, and uncovering new insights into devastating conditions such as cancer or rare genetic diseases. As scientists, we have a responsibility to help frame these developments, and participate in an open dialogue about what they mean for society as a whole.

As we celebrate EMBL's 50th and EMBL-EBI's 30th anniversaries in 2024, we have an opportunity to draw lessons from our history and look forward to exciting new developments just over the horizon.

We are extremely grateful to our funders, including EMBL's member states, UK Research and Innovation, Wellcome, the European Commission and the National Institutes of Health for their continued support which enables us to keep the world's biological data open and accessible to all.

## EMBL-EBI's Senior Management Team



**Ewan Birney**

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



**Jo McEntyre**

Deputy Director  
of EMBL-EBI



**Rolf Apweiler**

Associate Director  
of EMBL-EBI



**Rachel Curran**

Head of EMBL-EBI  
Administration & Operations  
and Deputy Chief Operating  
Officer of EMBL

# 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, and provides training and state-of-the-art research infrastructures. EMBL is an intergovernmental organisation with 29 member states, one associate member state, and one prospect member state. EMBL's latest 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.

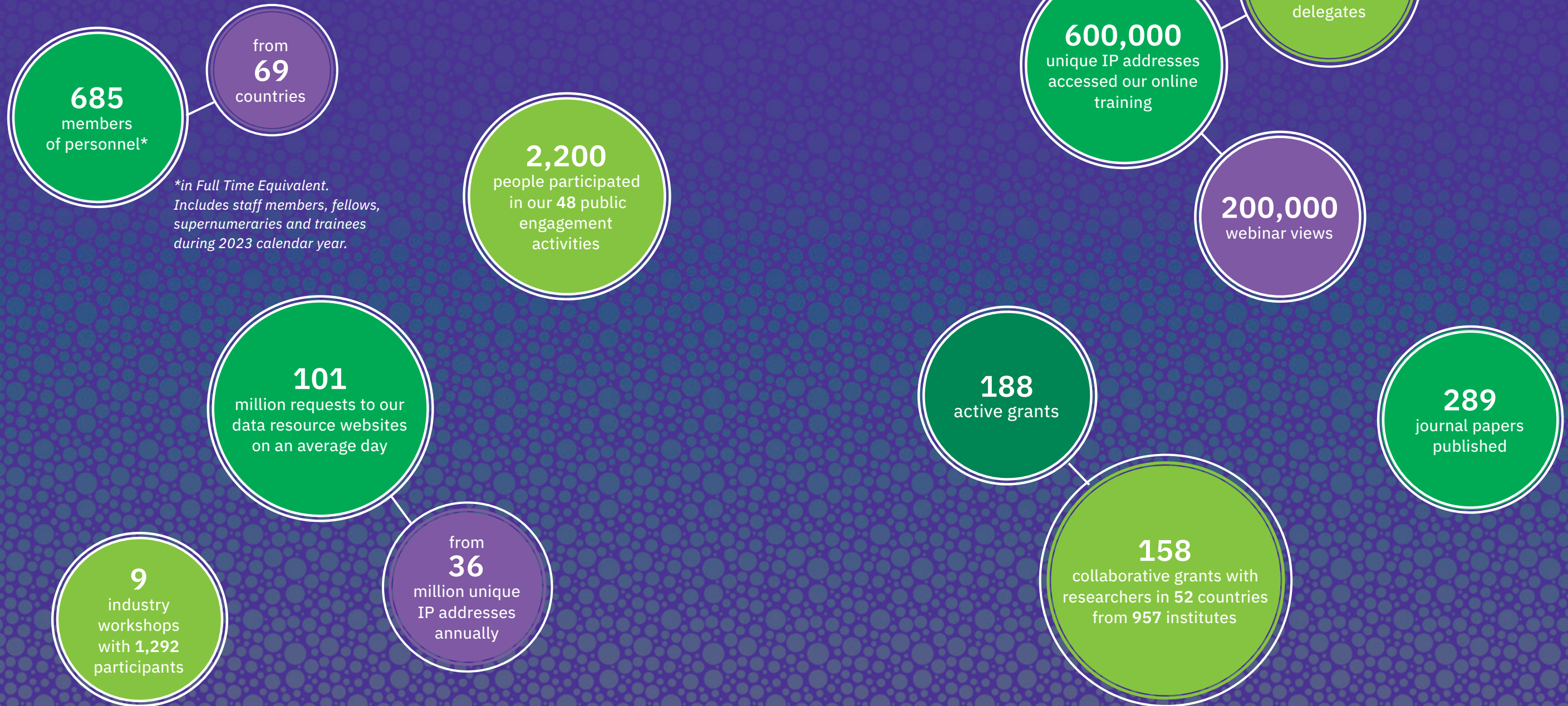
## Our missions

- 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 and applications of science.
- To support, as an ELIXIR Node, the coordination of biomolecular data provision in Europe.





# 2023 in numbers





# Celebrating 30 years of excellence in bioinformatics



*“As a hub of global excellence in bioinformatics, EMBL-EBI’s work is pivotal in harnessing the power of biological data for transformative scientific discoveries. By enhancing EMBL-EBI’s capabilities, we are ensuring that the global research community has the tools and resources necessary to meet rapidly evolving global challenges.”*

**Dame Ottoline Leyser, Chief Executive of UKRI**

*“EMBL-EBI is playing a central role in the emerging field of digital biology. Its world-leading data resources and technical expertise help nurture the virtuous circle of open science on which pioneering tools like AlphaFold can be built. By helping unlock the potential of bioinformatics, EMBL-EBI is accelerating discovery and enabling the scientific community to push the boundaries of our understanding of the world.”*

**Demis Hassabis, Co-Founder and CEO of Google DeepMind**



*“Since 2020, EMBL-EBI has seen a major increase in demand for data deposition and downloads, with one driver being the COVID-19 pandemic. Without open data, the real-time tracking of the virus and fast development of vaccines wouldn’t have been possible. With the launch of the AlphaFold database and the growth of our imaging archives, demand for our data resources today remains higher than ever before.”*

**Jo McEntyre, Deputy Director of EMBL-EBI**

*“In the FANTOM Consortium, we develop technologies and broadly use them to identify genome regulatory elements, promoters and enhancers, as well as RNAs for protein coding genes and non-protein coding RNAs. We count on EMBL-EBI to ultimately collect, elaborate and integrate data in the broader body of knowledge, making our data reusable by the whole community, and to ultimately support broad usage that will help mankind to cure diseases.”*

**Piero Carninci, Head of Genomics Research Centre, Human Technopole**



*“This year marks EMBL’s 50 anniversary and EMBL-EBI’s 30 anniversary, making it the perfect opportunity to celebrate EMBL’s pioneering role in the life sciences and EMBL-EBI’s contribution to making the world’s biological data openly accessible to all.”*

**Edith Heard, Director General of EMBL**

*“EMBL-EBI has been instrumental for SciLifeLab, providing access to an extensive range of data and analytical tools spanning gene sequences and protein structures. As we embark on an era characterised by an increasingly data- and AI-driven scientific exploration, the potential for EMBL-EBI’s resources to be leveraged for groundbreaking research supporting better health for people and the planet grows exponentially.”*

**Olli Kallioniemi, Director of SciLifeLab**



*“As new technologies make the life sciences more data-driven, scientific careers require grounding in more than one STEM subject, and in particular in data management and analytics. Training is key for creating a global workforce that develops solutions to address urgent challenges.”*

**Cath Brooksbank, EMBL-EBI Head of Training**

*“Rapid advances in AI are accelerating structural biology research, enabling a better understanding of biological systems at the molecular level. As the keeper of the world’s biological data, EMBL-EBI is uniquely placed to leverage and support new developments in AI and maximise its impact by encouraging responsible and prudent use of the technology.”*

**Sameer Velankar,  
EMBL-EBI Team Leader for the Protein Data Bank in Europe**



*“EMBL-EBI’s diverse and international community thrives on collaboration. In the future, we’re keen to create even more opportunities for knowledge sharing and cooperation, bringing different backgrounds and skills together to keep the world’s biomolecular data flowing.”*

**Rachel Curran, EMBL Deputy Chief Operating Officer & Head of EMBL-EBI Administration & Operations**



# 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, chemistry, literature data and more.

## World's leading bioinformatics institute

In 2023, EMBL-EBI's Scientific Advisory Council, consisting of a panel of experts from around the world, reviewed the institute's data resources. They resoundingly endorsed the data resources and were extremely complimentary about how EMBL-EBI had delivered them, especially in the context of the COVID-19 pandemic.

*“Over the last three years, EMBL-EBI has arguably risen to be the leading bioinformatics service and research institute worldwide.”*  
– 2023 SAC Review

## The global gateway for pathogen data

Building on its COVID-19 Data Portal, EMBL-EBI launched the [Pathogens Portal](#) – an online platform that enables researchers, clinicians and policymakers to access the world's most comprehensive collection of biomolecular data about pathogens.

The portal spans over 200,000 pathogen species and strains, and is set to become a key tool for infection biology and pathogen surveillance. The ability to share pathogen data across borders is crucial, especially during public health outbreaks and pandemics.

The resource enables access to nucleotide sequences, raw genomic data, sample metadata, and relevant scientific literature, with more data types to be integrated in the future. It also features a cohort browser, which contains highly sought-after clinical-epidemiological data from patient cohorts.

*“The Pathogens Portal is an important step in preparing for the next pandemic.”*

**Marion Koopmans,**  
Head of the Erasmus Medical Centre's  
Department of Viroscience

## Rare disease database joins EMBL-EBI

Genomic variants linked to rare diseases can be difficult to interpret, and a diagnosis can take years. The recurring tests and delays can take a heavy toll on patients and families.

By sharing rare disease data, clinicians can identify the genotype-phenotype correlations underpinning rare diseases, and assist diagnosis. The [DECIPHER](#) database, founded in 2004 at the Wellcome Sanger Institute, is used by researchers and clinicians to interpret and share phenotypic and genotypic data from 46,000 rare disease patients around the world.

DECIPHER has now moved to EMBL-EBI, to further strengthen its links to the wider bioinformatics ecosystem and benefit from the institute's data management expertise.



*“When I'm reviewing patients in the clinic, DECIPHER gives me a snapshot of what is known about a variant, enabling me to quickly assess whether it explains a patient's symptoms, and whether new knowledge has emerged since it was reported. This is crucial to ensure patients receive the most up-to-date, high-quality advice in such a fast-moving field.”*  
**Helen Firth, Consultant Clinical Geneticist and Hon. Professor of Clinical Genomics, Cambridge University Hospitals**



## A more diverse human reference genome

On average, human genomes are more than 99% identical. However, the small differences contribute to each person's uniqueness, and can provide insights about their health, helping to diagnose disease and guide medical treatments.

These differences also mean that using one standard reference genome, as many studies currently do, has limitations. To address these limitations, the Human Pangenome Reference Consortium, which includes EMBL-EBI, has released a high-quality collection of reference [human genome](#) sequences called the human pangenome. This captures diversity from different human populations better, compared to the current human reference genome.

The first draft of the human pangenome includes the maternal and paternal genome sequences from 47 people, with the aim of increasing this number to 350 people. As part of the project, EMBL-EBI's Ensembl team painstakingly mapped the high-quality annotations for the reference human genome across the pangenome. The human pangenome data are now openly available to all, in the Ensembl database, and will be updated as the project progresses.

## New tools for drug discovery

Understanding which genetic mutations cause disease is crucial for drug development. Missense variation is a type of mutation in which a single nucleotide change results in a different amino acid within a protein.

Understanding how missense variation translates into proteins is vital for identifying drug targets and developing new therapeutics, but it requires the assimilation of a broad range of data. Several tools now exist to explore this research area.

[ProtVar](#), developed by EMBL-EBI, helps researchers contextualise and interpret human missense variation within proteins. Clinical geneticists can use this information to identify and prioritise which genetic variants are most likely associated with disease.

Similarly, Google DeepMind's AlphaMissense is a new AI model which classifies missense variants. EMBL-EBI [has integrated AlphaMissense](#) into the Ensembl's Variant Effect Predictor, a powerful tool that allows users to input a list of genetic variants and predicts which genes or regulatory elements they fall within, and what impact they may have.



## Case study: How EMBL-EBI supports genomic medicine

### The challenge

In the clinic, genome sequencing can support healthcare by enabling quicker diagnosis, matching patients to treatments, and monitoring response. Genomic medicine is particularly impactful for rare disease and cancer patients.

But in order to interpret genomic sequencing results, clinicians need reference data – for example, to work out whether the version of a gene that a patient is carrying is common, or a rare mutation.

In cancer, for example, each patient has a unique constellation of mutations, so gleaning insights from genomic sequencing data requires sophisticated analysis tools. EMBL-EBI researchers develop such tools to understand the causes and progression of cancer.

Finally, genomic medicine requires a skilled workforce able to build clinical bioinformatics workflows and interpret results. EMBL-EBI provides a wide range of genomics training materials.

### Where EMBL-EBI comes in: Reference data, analysis tools and training

Patients in many countries are already seeing benefits from genomic medicine. The UK's National Health Service (NHS) is the first to offer whole genome sequencing as part of routine care, via a partnership with Genomics England.

Genomics England workflows depend on EMBL-EBI's data resources, and use these reference data in bulk. Each month, thousands of patients receive genome sequencing results which rely on EMBL-EBI data and tools.

Every time the NHS analyses sequencing data to return a diagnostic report to a patient, it uses EMBL-EBI's Ensembl gene annotation database, and tools like Ensembl's Variant Effect Predictor, which determines the effect of changes to the genetic code.



*"Without the reference genomic data managed by EMBL-EBI, it would be impossible to interpret patients' genomic tests. It would be like trying to solve a puzzle with billions of pieces from slightly different versions of the same jigsaw, and no guiding picture."*

**Ewan Birney, Director, EMBL-EBI and Deputy Director General, EMBL**





## Empowering proteomics research with machine learning

Preparing proteomics data for machine learning is complex and time-consuming. Different labs use varied methods, making it hard to share and use data widely. ProteomicsML is a free online resource that tackles this challenge by offering an online platform with easy-to-use data formats and detailed tutorials to aid accessibility across the field.

ProteomicsML also facilitates the application of machine learning to proteomics data by offering openly available datasets that can be used to train machine learning algorithms, and providing educational materials to help researchers get the most out of these datasets.

*“The aim of the platform is to promote AI and machine learning applications for mass spectrometry-based proteomics data. The community will create and document training datasets and tutorials, making it a vital resource for anyone new to the field or looking to apply machine learning to proteomics data.”*

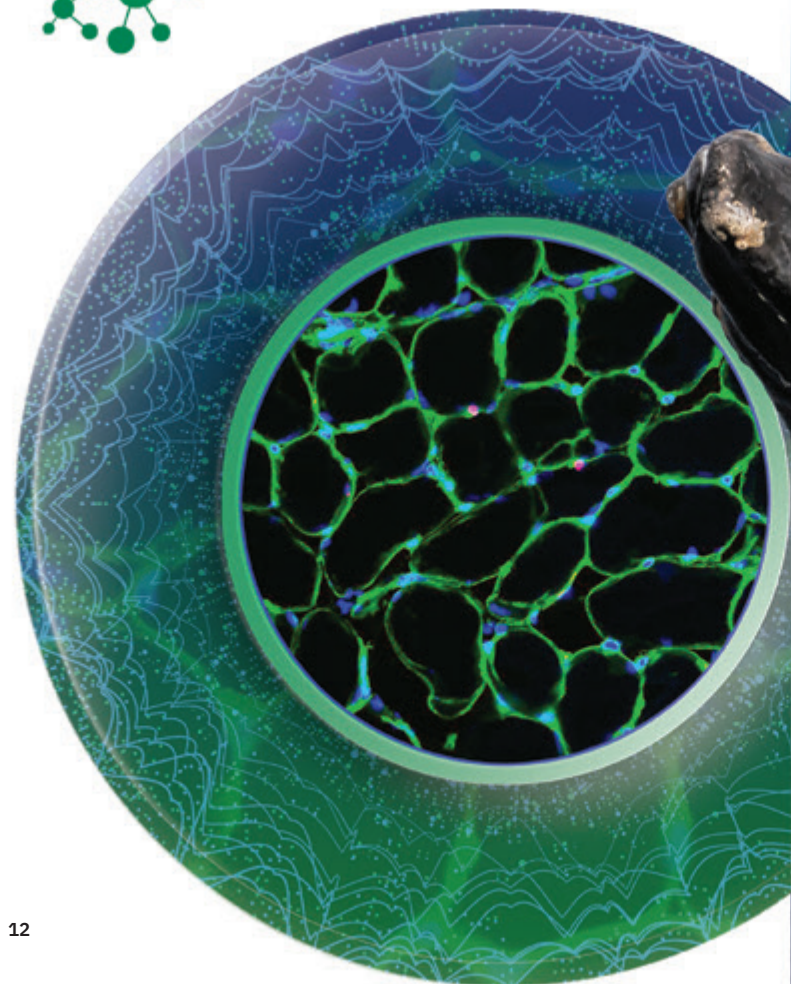
**Juan Antonio Vizcaino,**  
Team Leader at EMBL-EBI

## Democratising access to microscopy data

Due to the wide range of microscopy technologies available, the bioimage landscape is very fragmented, with data being produced in over 50 file formats. Comparing data across formats is a major challenge.

To address this obstacle, EMBL-EBI’s BioImage Archive team has converted different formats into one community-driven, open source, file format called OME-Zarr. The implementation of the new file format has enabled the development of [visual galleries in the BioImage Archive](#) – collections of frequently accessed images that users can explore in their browser, without downloading them.

The new galleries could be particularly valuable for users who do not have access to significant storage space or a robust internet connection. There is also a dedicated AI gallery, especially developed for training and testing artificial intelligence (AI) and machine learning tools.



## Enabling biodiversity research worldwide

Sequencing and storing the genomic data of all species on Earth is crucial for conservation and biodiversity research. These data enable researchers to understand the genetic factors that make a species unique, including its vulnerabilities and strengths in adapting to changing environmental conditions. The information also has direct applications in agriculture and bioengineering. Understanding plant genomes supports the development of more resilient crops and more efficient biofuels.

To enable biodiversity researchers to leverage genomic data, EMBL-EBI has launched the [Global Biodiversity Portal](#) – a centralised, open-access database that consolidates genomic information collected by the Earth BioGenome Project.

The portal offers user-friendly search filters and features, including sequencing status tracking. Designed for scalability and regular updates, the portal harmonises data from diverse sources, and keeps information current.

*“This portal represents a significant advancement for the scientific community, providing a comprehensive platform for accessing and sharing vital genomic data.”*

**Camila Mazzoni, Group Leader,**  
**Leibniz Institute for Zoo and Wildlife Research Berlin**





## Supporting the search for sustainable animal feeds

Meat production is a very resource-intensive industry. Among many other challenges, the feed that animals consume needs to be nutritious and sustainable.

The HoloFood project, funded by Horizon 2020, has been exploring the relationship between farmed animals, the food they eat, and the microbes living in their guts, focusing on chicken and salmon in particular. The aim of the project is to support the development of more sustainable animal feeds.

After four years of data collection, with the help of EMBL-EBI, the project has launched the [HoloFood Data Portal](#). This is the world's first freely accessible database containing comprehensive biomolecular data that enables researchers to understand the microbial diversity in the chicken and salmon gut, and how changes in diet affect animal health.



*"We have shown the potential of using the gut microbiome to investigate more sustainable and cost-effective animal feeds, and we hope other researchers and companies will continue this work."*

**Varsha Kale, Senior Bioinformatician, EMBL-EBI**

### TIMOTHÉ CEZARD

#### PROJECT LEAD - EUROPEAN VARIATION ARCHIVE

Originally from Paris, Timothé has shifted his focus from biology, to bioinformatics, software development and now infrastructure building. An avid user of EMBL-EBI's data resources, he was keen to "look under the bonnet" and use his expertise to help develop them.

The European Variation Archive (EVA) enables scientists to decipher genomic variation in all species, which is key for understanding human health, animal breeding and agriculture, or fighting pathogens. Timothé and his team are currently working on improving the EVA's data submission process, and integrating variation data in the wider ecosystem, to give users a more complete picture of their variants of interest.

Asked what bioinformatics will look like in the future, Timothé said: *"Currently the world is generating huge volumes of genetic variation data in many key species. EMBL-EBI is at the centre of ensuring this data is available to all scientists to enable future scientific discoveries."*



## Research

EMBL-EBI's research groups use computational methods and machine learning to make sense of vast, complex datasets. Our researchers work closely with experimental scientists, increasingly tackling problems of direct significance to medicine and the environment.

### Shedding light on cancer immunotherapy

Research from the Cortes-Ciriano group and collaborators at MIT and Cold Spring Harbor Laboratory has shed light on [why immunotherapy does not always work](#) in certain types of cancer.

Immune checkpoint blockade (ICB) is an approved immunotherapy for cancer treatment, which harnesses the power of the patient's immune system to target and destroy cancer cells. However, ICB is only effective in some patients and it is still unclear why.

This study looks at the complex mechanisms underlying response to ICB in tumours with DNA mismatch repair deficiency. Although these tumours have a relatively high response rate to ICB, most patients do not respond, which poses significant challenges in optimising treatment.

The researchers used mouse models and cell lines to understand what is happening at a molecular level, and demonstrated that inactivation of the DNA mismatch repair pathway is not enough to improve patient responsiveness to ICB.



*"This is an important body of work that provides new insights into the factors that control immune responses against cancer and why some tumours fail to respond to immune-stimulating therapies."*

**Tyler Jacks, Professor at the Koch Institute at MIT**



## Streamlining dynamic nanopore sequencing

Long-read nanopore sequencing has revolutionised the way scientists obtain genomic data due to its speed, accessibility and portability. But like any new technology, it requires the development of new data analysis tools.

BOSS-RUNS is a new open source method developed by the Goldman group and colleagues at the University of Nottingham.

BOSS-RUNS helps scientists to dynamically adapt their nanopore sequencing runs to make the process faster and more efficient. Similarly, BOSS-RUNS is ideal to use when analysing multiple genomes in the same sample, for example in a microbiome. BOSS-RUNS is a shining example of useful tool development for genome sequencing.



## Understanding the mechanisms of enzyme catalysis

Enzymes are important to life because their genes constitute a large fraction of genomes – 20% in humans. Despite decades of studies and a rich literature dedicated to understanding the reaction mechanisms of enzymes, the rules of enzyme catalysis are still not entirely clear.

A new tool called [EzMechanism](#), developed by the Thornton group has made progress towards predicting the mechanisms of enzymes that have not been characterised yet. EzMechanism aims to answer the question ‘given an enzyme whose reaction and structure we know, can we predict its mechanism?’



*“Enzymes perform the chemistry of life, so understanding what functions they fulfil and how they do it makes for fascinating research and has numerous practical applications.”*

**Dame Janet Thornton,**  
Director Emeritus at EMBL-EBI

## New insights into rare eye disorders

The retina is found at the back of the eye. It’s a layered tissue that converts light into a signal that can be interpreted by the brain. Each retinal layer comprises different cell types that play a unique role in this light conversion process.

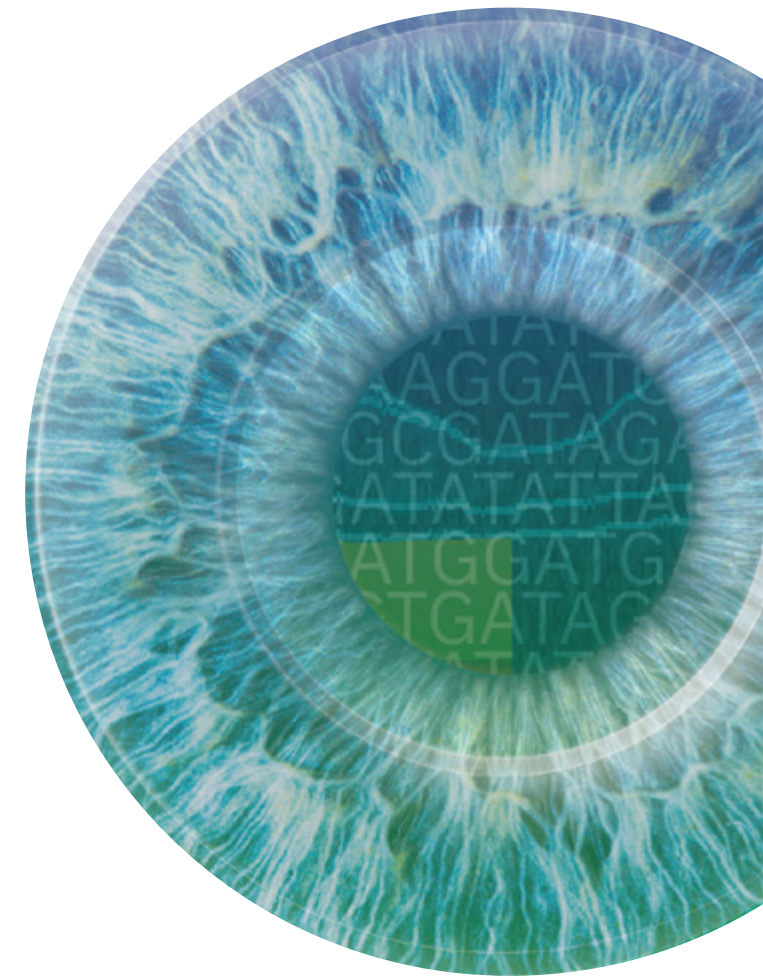
Researchers in the Birney group have analysed imaging and genomic data from the UK Biobank to uncover [new insights](#) into rare diseases of the human eye. These include retinal dystrophies – a group of inherited disorders that are the leading cause of blindness certification in working-age adults.



*“Systematic bioinformatic analysis of large-scale participant data cohorts is driving the future of genomic medicine.”*

**Omar Mahroo, Professor of Retinal Neuroscience at University College London and Consultant Ophthalmologist at Moorfields Eye Hospital.**

The work identified genomic variations associated with the thickness of one or more of the photoreceptor cell layers in the eye. The newly identified genomic associations can be openly accessed through EMBL-EBI’s GWAS Catalog.





# Training and industry partnerships

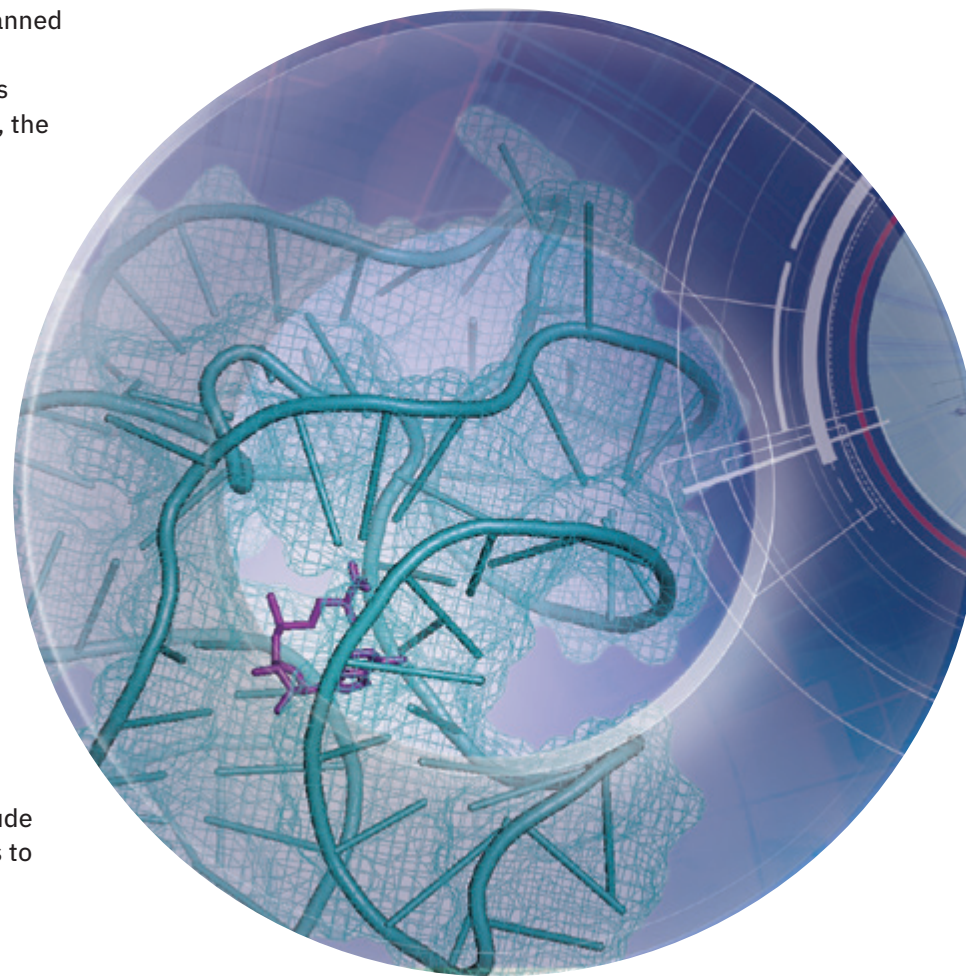
EMBL-EBI delivers world-leading training in biodata science including at predoctoral and postdoctoral levels. We empower scientists at all career stages to make the most of biological data, and to strengthen bioinformatics capacity across the globe.

In 2023, EMBL-EBI's Training programme spanned 21 live courses with 560 delegates, and 37 webinars. With more than 600,000 unique IPs accessing our impressive online training offer, the programme's global reach continues to grow.

## World-leading PhD and postdoctoral programmes

Characterised by first-class training, internationality, dedicated mentoring and early independence in research, the EMBL International PhD Programme is among the world's most competitive PhD training schemes.

EMBL-EBI offers several postdoctoral programmes in a range of bioinformatics and related areas. The programme strengths include interdisciplinarity, internationality and access to world-class facilities.



### JACQUELINE TOUSSAINT

#### PREDOCTORAL FELLOW

After studying biochemistry and microbiology at Montana State University, Jacqueline took a deep dive into the world of bioinformatics by joining EMBL's international PhD programme, as part of the Lees group at EMBL-EBI.

Jacqueline's work aims to develop tools to understand bacterial genetics, like how genotypes correlate with phenotypes, or traits. She is particularly interested in antimicrobial resistance and virulence – a microorganism's ability to cause damage to a host. This is driven by her curiosity about "the strange world of bacteria" and by a desire to improve healthcare.

Musing on what bioinformatics would look like in the future, Jacqueline said: *"I think the scale at which data is going to be produced, stored and processed will be beyond our comprehension. It may also get easier to do bioinformatics. In the past, bioinformatics software development had a high barrier to entry, but it could become as simple as plug 'n' play."*



### IGUARACY PINHEIRO DE SOUSA

#### POSTDOCTORAL FELLOW

While studying genetics and molecular biology in Brazil, Iguaracy developed an interest in bioinformatics. His postdoctoral research in the Petsalaki group focuses on leveraging the increasing volumes of single-cell transcriptomics data to address some of the limitations of organoids. To do so, he has set up international collaborations with wet lab colleagues, specifically looking at the heart and brain.

Iguaracy is a strong advocate for collaborative science, and finds his inspiration by chatting to colleagues from different areas. One of his more futuristic projects focuses on precision medicine, and exploring if blood vessels could be used as 'nanobot highways' to reach infection sites in the body.

Asked what bioinformatics will look like in the future, Iguaracy said *"As AI advances, the focus of bioinformaticians is likely to change from coding and programming tasks to more interpretive roles, and problem-solving."*





## Preparing for a career in research infrastructures

ARISE is EMBL's career accelerator for people who want to work in research infrastructures. This unique three-year programme is the perfect opportunity to work on method and technology development, get bespoke career support and access to exclusive training curricula.



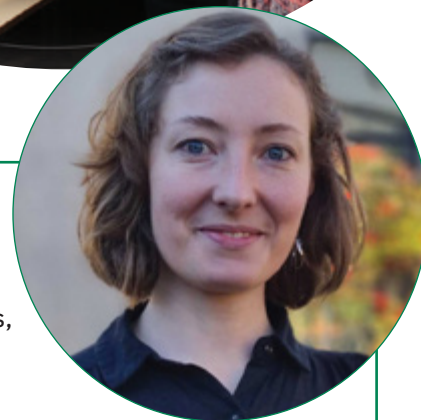
### MELANIE SCHNEIDER

#### ARISE FELLOW

Melanie Schneider first joined EMBL-EBI as a Computational Scientist working on understanding targets for drug development. During this project, Melanie developed an interest in how research infrastructures, such as those delivered by EMBL-EBI, support the work of scientists worldwide.

Melanie took a leap of faith to pursue this new interest by applying for the [ARISE Programme](#), EMBL's career accelerator for scientists working in research infrastructures. This has been an opportunity for Melanie to learn new skills, including method development, user support and strategic management. Her current project brings together bioinformatics expertise from EMBL-EBI and structural biology knowledge from EMBL Grenoble.

*"The programme organisers are very open to feedback and considerate of the fellows' needs," explained Melanie. "For example, when I went on maternity leave, they were incredibly supportive, and helped me organise things so I didn't feel like I was missing out."*



## Tackling the global food security challenge

EMBL-EBI is responding to global food security challenges by working with industry and academic partners to optimise the utility and value of publicly-available biological data.

The Agri-Tech Partnership brings together agricultural technology (agri-tech) domain experts to help us build a fit-for-purpose EMBL-EBI AgriData Platform integrating EMBL-EBI data and other publicly available data worldwide. The platform includes computational tools to extract and interpret knowledge for real-world applications.

The consortium is looking for organisations and individuals that will:

- Share their use cases and determine which datasets to prioritise
- Flag gaps in existing data and tools
- Collaborate on projects to close gaps or to generate new data

The aim is to develop and build data portals, tools and services that are appropriate for commercial research and development, in areas such as genetic crop improvement, crop protection products, environmental safety protection, with the ultimate goal of increasing food supply.



*"The wealth of existing bioinformatics knowledge and expertise in the biomedical domain ideally positions EMBL-EBI to translate these skills into addressing the bioinformatics challenges of the agricultural domain."*

**Effie Mutasa-Gottgens,**  
Senior Scientific Officer



Read more





# Culture and infrastructure

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

## Infrastructure upgrade

Since the COVID-19 pandemic, EMBL-EBI has seen a steep increase in the volumes of data that researchers deposit in its data archives, as well as a doubling of the number of researchers using its services. The growth of new technologies such as genomic sequencing, bioimaging and artificial intelligence (AI) are driving this increased data production and usage.

In response to this increase, UK Research and Innovation (UKRI) confirmed £80.7 million of funding for EMBL-EBI. The funding comes from the UKRI Infrastructure Fund and will be distributed over the course of six years. It is set to help EMBL-EBI transform its technical infrastructure to meet the data needs of the life sciences community.

*“Since its inception 30 years ago, EMBL-EBI has become a foundational global research infrastructure. The data resources, tools and services developed by EMBL-EBI are essential for researchers in delivering bio-based solutions for the challenges we face as a society.*”

*We are pleased to continue to support EMBL-EBI in its mission to make research data freely and openly available to the world. This enables scientific discoveries in all areas of bioscience, from food security to infectious disease, genomic medicine and biodiversity conservation.”*

**Dr Amanda Collis, Interim Deputy Executive Chair at the Biotechnology and Biological Sciences Research Council (BBSRC), part of UKRI**



## Thornton building progress

Construction work on EMBL-EBI’s third permanent building on the Wellcome Genome Campus, the Thornton Building, progressed significantly in 2023. The building – designed with sustainability and collaboration in mind – will be a dedicated space for innovation and translation.

Serving as a new home for the Open Targets public-private partnership, which uses human genetics and genomics data for systematic drug target identification and prioritisation, the Thornton building also features dedicated spaces for close collaborations with industry. The Thornton building is expected to officially open in 2025.



### ALANNAH LOCKWOOD

#### PRINCIPAL LEGAL OFFICER

After an earlier career in advertising and marketing law, Alannah worked for several international non-governmental and non-profit organisations before joining EMBL-EBI. Much of the work she and her team do supports external funding and procurement contracts. This can be a complex process for inter-governmental organisations like EMBL, which have a special legal status under international law.

*“Reviewing contracts can become almost routine, but at EMBL you have to retrain your brain to spot things that are not compatible with the organisation’s special status, privileges and immunities.”*

Alannah sees this challenge as an opportunity to come up with creative approaches and negotiate outcomes that work for all parties. Her role also involves building relationships with funders, collaborators and suppliers. This is an essential part of the role that can open up new funding and partnerships to help advance EMBL-EBI’s science and technology.

*“One project we support with EMBLEM colleagues is the AlphaFold Database developed by EMBL-EBI and DeepMind. It’s been an interesting project and it’s been great to see its impact since launch.”*





## Public engagement: data science for all

EMBL-EBI focuses on building long-lasting relationships with community groups that have little exposure to data science. To do this, we work with public groups to develop activities that bring people together, to listen to and learn from each other while exploring bioinformatics and its impact.

Working with community partners EMBL-EBI delivered a range of activities in Cambridgeshire, Suffolk and Norfolk.

Our partnership with Suffolk Family Carers is an example of how our engagement partnerships work in practice. EMBL-EBI connected with Suffolk Family Carers during an audience scoping project in 2021 and have been working with the Young Carers team since then.

Using images from the [BioImage Archive](#) as inspiration for writing science haikus, EMBL-EBI developed and delivered a creative writing activity for 11-16 year olds. The activity was formed of two virtual workshops and one in-person extended session in a local town community space.

EMBL-EBI also welcomed two groups of young people to the Wellcome Genome Campus for a day packed with DNA themed activities, including DNA extraction, exploring RNA, and the [wetlands nature trail](#). The older age group (13-17 year olds) had an opportunity to chat to EMBL-EBI colleagues about their roles, career paths and the day-to-day reality of working in science and technology.

*"I don't usually like science at school, but really enjoyed learning about DNA and RNA and being in a different setting."*

**Participant**



Watch our video



## Making bioinformatics GREENER

As science increasingly relies on big data and AI tools, the carbon footprint of computational work is on the rise. To address the issue, EMBL-EBI researchers and collaborators have put forward the [GREENER principles](#) for how the computational science community could make its practices more environmentally sustainable.



*"We all know that our travel and diet choices affect our carbon footprint. As scientists, we should also be mindful of the silent and unintended environmental consequences of our computational work. The GREENER principles are a good starting point for understanding and mitigating this impact."*

**Loïc Lannelongue, Research Associate, University of Cambridge**

## Reducing energy consumption while increasing efficiency

Following a sustained, multi-year effort, EMBL-EBI's technical teams have successfully reduced the energy consumption of its world-leading technical infrastructure, while improving efficiency.

The transformation included the development and implementation of a tiered storage system for the vast volumes of data that EMBL-EBI manages. This enabled a reduction of approximately 8% in data centre electricity usage, while also increasing storage and compute capacity.

Future-proofing technical infrastructure in this way is enabling EMBL-EBI to adapt to the increasing volumes of data depositions it receives, and to deliver world-class bioinformatics services long into the future.



**MOHAMED ALIBI****HIGH PERFORMANCE COMPUTING ENGINEER**

Mo studied computer networking at the University of Sciences in Tunisia, and joined EMBL-EBI to support the technical side of the institute's comprehensive training programme. To further hone his skills, he transitioned to the role of High Performance Computing (HPC) Engineer.

His work supports the technical infrastructure that enables the fast and seamless data annotation done by EMBL-EBI experts before datasets are made publicly available. This adds value to the data generated by the global scientific community, making it more Findable, Accessible, Interoperable and Reusable (FAIR). HPC is energy-intensive, but one of the key priorities for Mo and his team is exploring new ways of reducing the carbon footprint of this critical process.

Asked what bioinformatics will look like in the future, Mo said: *"It's going to be crazy, because we're only starting to scratch the surface of what can be done with machine learning and AI."*



## An open, inclusive and international community

EMBL-EBI's staff and fellows come from almost 70 countries, bringing together an impressive range of backgrounds and expertise. They are the institute's strongest asset and create a culture of openness and knowledge exchange.

A survey of our staff and students showed a strong sense of community within teams and more widely, within the institute, as well as pride in working at EMBL-EBI. The top words used by staff to describe the organisational culture were "open", "inclusive" and "friendly".

# New leadership

Meet the team leaders who joined EMBL-EBI in 2023.

**David Hulcoop**

David Hulcoop was promoted to Executive Director of Open Targets, the public-private partnership which uses human genetics and genomics data for systematic drug target identification and prioritisation. David was previously the Strategy and Operations Director at Open Targets.

**Manu Morgan**

Manu Morgan has been appointed as Deputy Head of Human Resources at EMBL-EBI. Manu will drive the development of a comprehensive career development framework and encourage an organisational culture of continuous learning and growth.

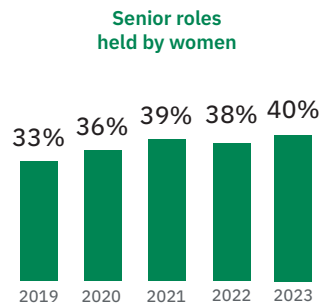
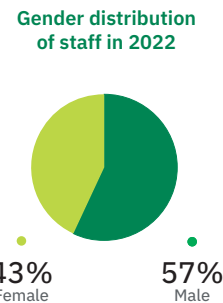
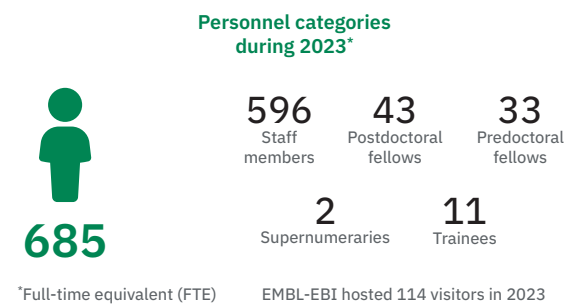
**Kyle Morris**

Kyle Morris has been appointed as a new Team Leader for the Electron Microscopy Data Bank, EMBL-EBI's public repository for electron cryo-microscopy maps and tomograms of macromolecular complexes and subcellular structures.



# Facts and figures

## Staff numbers



\* These numbers reflect staffing levels during the entire 2023 calendar year.

## Financial information

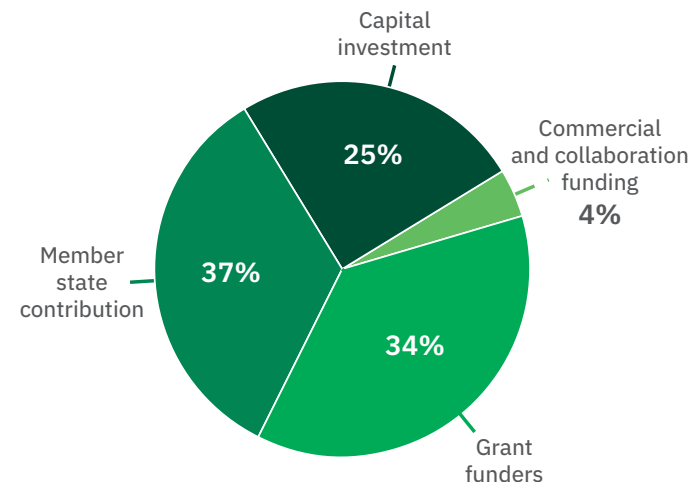
We are grateful to EMBL member states and other funding bodies for their continued support. They enable us to maintain and grow our data resources,

conduct vital research and training, and respond to the changing requirements of the international scientific community.

## Funding sources

EMBL-EBI's financial year runs from 1 January to 31 December.

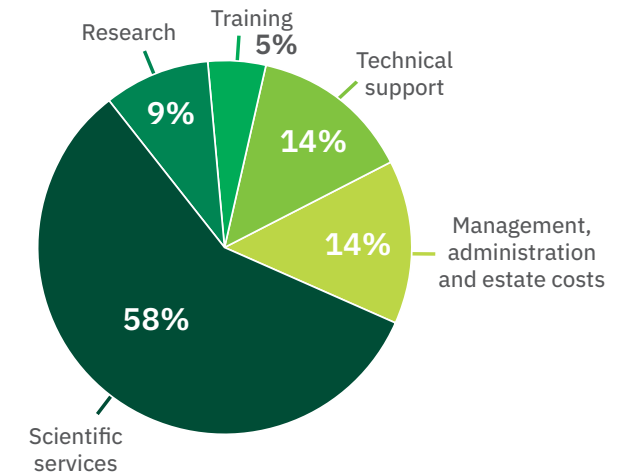
EMBL-EBI receives its funding through four main funding channels. The operating expenditure is funded by EMBL member state contributions (37%), commercial and collaboration funding (4%) or external funding bodies via grant awards (34%). EMBL-EBI also receives significant capital awards for investment in buildings and data infrastructure (25%).



## Operating expenditure

The total operating expenditure of EMBL-EBI in 2023 was €98.8m from all funding sources.

Scientific services accounted for 58% of the expenditure in support of EMBL-EBI's data resources, with a further 9% 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 14% on management, administration, and estate costs.



## Capital investment

EMBL-EBI's Data Infrastructure for Life Sciences programme (2019 - 2024) oversees £70 million of capital investment, of which £45 million comes from the UK government's Strategic Priorities Fund via UK Research and Innovation (UKRI). This enables the transformation of storage, compute and networking facilities as the demand for our data resources continues to grow, and helps to establish the BioImage Archive, a central, open data resource for biological images.

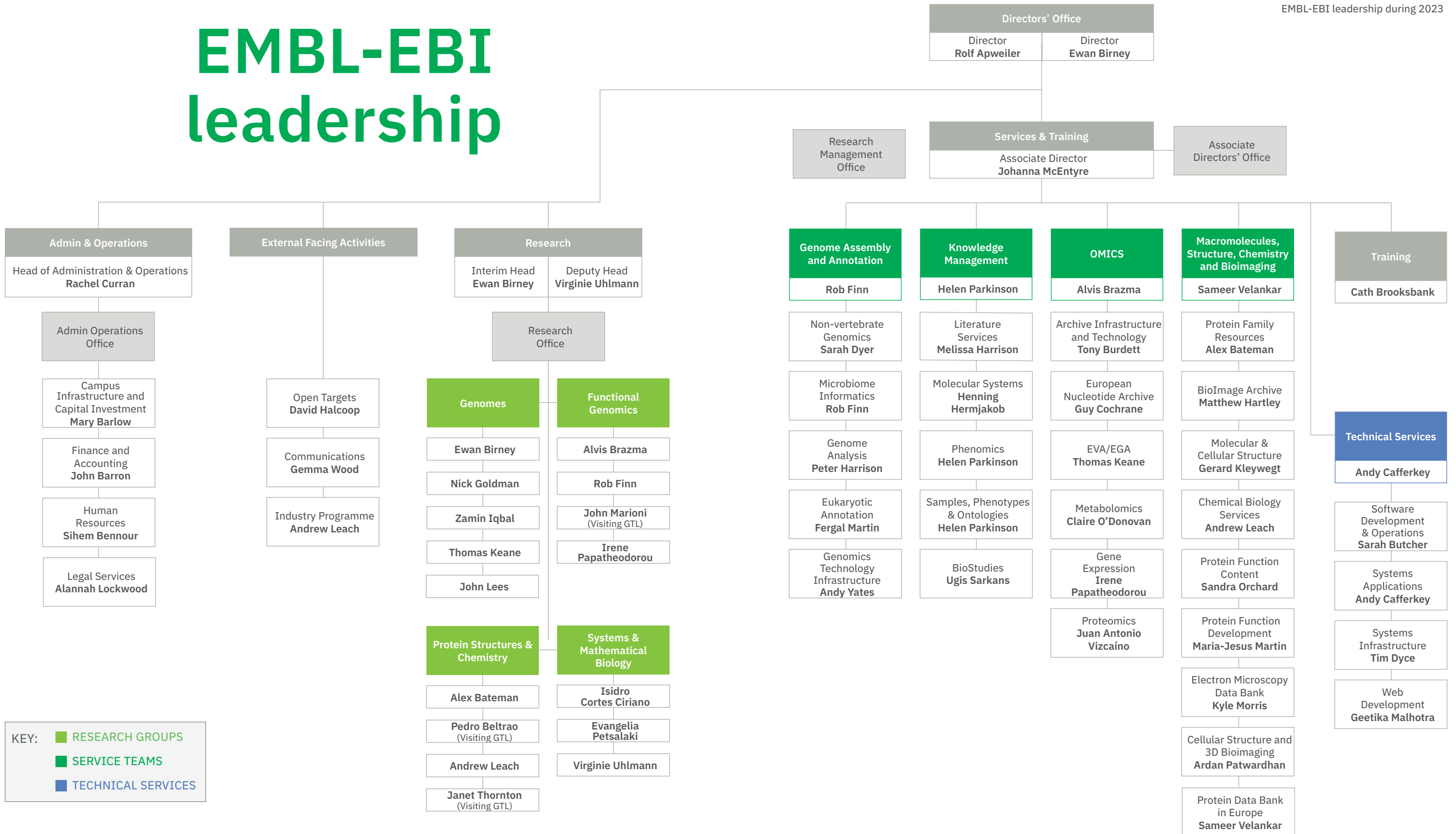
The programme supports EMBL-EBI's Thornton Building, the institute's third permanent building on the Wellcome Genome Campus. The building is set to complete in 2024 and was funded by UK Research and Innovation (UKRI), Wellcome and the Biotechnology and Biological Sciences Research Council (BBSRC). It is named after the institute's Director Emeritus and esteemed scientist Professor Dame Janet Thornton.

## Strategic partners and funders

EMBL-EBI works closely with strategic partners and funders from across the globe. Alongside funding from EMBL member states, in 2023 the institute received €43.3m in research and service grant funding from 25 funders providing support for 188 projects.



# EMBL-EBI leadership



KEY:

- RESEARCH GROUPS
- SERVICE TEAMS
- TECHNICAL SERVICES



# 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 one prospect member state. 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 2023, EMBL-EBI was led by joint Directors Rolf Apweiler and Ewan Birney, with the support of Associate Director of Services, Johanna McEntyre and the Head of Administration and Operations, Rachel Curran. In early 2024, Johanna McEntyre was appointed Deputy Director of EMBL-EBI, and Rolf Apweiler became Associate Director of EMBL-EBI.

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

Chan Zuckerberg Initiative

Cancer Research UK

Connective Tissue Oncology Society

European Association for Cancer Research

European Commission

Engineering and Physical Sciences Research  
Council

Economic and Social Research Council

Fonds National de la Recherche

The Leona M. and Harry S. Helmsley  
Charitable Trust

The Mark Foundation for Cancer Research

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 UK

Schmidt Futures

The Pezcoller Foundation

University of Cambridge

UK Research and Innovation

Wellcome

## Imprint

### Publisher

EMBL-EBI

### Portrait image credits

p. 6 (Dame Ottoline Leyser): UKRI

p. 6 (Demis Hassabis): Google DeepMind

p. 6 (Piero Carninci): Human Technopole

p. 7 (Edith Heard): Kinga Lubowiecka/EMBL

p. 7 (Olli Kallioniemi): SciLifeLab

p. 9 (Helen Firth): Wellcome Sanger Institute

p. 15 (Tyler Jacks): Steve Boxall

p. 17 (Omar Mahroo): Omar Mahroo

p. 20 (Melanie Schneider): Mylène André

p. 22 (Dr Amanda Collis): UKRI

p. 25 (Loïc Lannelongue): Loïc Lannelongue

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

### Other image credits

p. 3 (EMBL-EBI South building): Jeff Dowling/EMBL

p. 20 (Advanced Light Microscopy Facility): Kinga Lubowiecka/EMBL

p. 24 (Public engagement activity): Jeff Dowling/EMBL

### Printer


Healeys. [healeys-printers.co.uk](https://healeys-printers.co.uk)



## EMBL's European Bioinformatics Institute (EMBL-EBI)

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

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

 +44 (0)1223 494 444

 [comms@ebi.ac.uk](mailto:comms@ebi.ac.uk)

 @emblemibi

 /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](http://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:** Serbia