10. Scientific Services

Introduction

EMBL's scientific services aim to provide world-class support to research communities throughout EMBL's member states and around the world. Whether experimental or data driven, they span many biological techniques and multiple domains within molecular biology. EMBL's scientific services include access to structural biology infrastructure, imaging services, cutting-edge genomics, proteomics, and other scientific core facilities, and wide-ranging molecular data resources for scientists at EMBL member state institutes and beyond. In this way, research questions can be addressed using interdisciplinary expertise that is largely unavailable elsewhere in Europe. **Throughout the next Programme, EMBL will strengthen and expand its services, both individually and by integrating new complementary services.** The aim is to accelerate European and global science by contributing to the fundamental research that is integral to solving solutions to societal challenges.

Accessible infrastructure, invaluable expertise, and quality assurance are the key principles that make up EMBL's scientific service provision. With these principles, EMBL aims to support and advance research and technology development at its six sites, within its member states, and globally. This distributes advances in scientific knowledge and technological developments across national and international networks to enrich scientific communities around the world.

1. **Accessible infrastructure**. In alignment with EMBL's dedication to open science, EMBL aims to provide access to complex, expensive, or large infrastructures to researchers across EMBL member states and beyond.
2. **Invaluable expertise**. EMBL scientists bring their invaluable and deep expertise in service infrastructure to researchers around the world. This can take the form of support provision via helpdesks, collaborative engagement to solve targeted experimental problems, or the provision of advanced training in complex technologies.
3. **Quality assurance**. Rigorous establishment, adherence to, and sharing of quality control and validation measures are essential to EMBL's scientific service provision. These ensure that experimental and data standards are set in collaboration with the community, paving the way for high-quality and consistent data production, data management, and access.

Alongside these principles, experts within EMBL's scientific services provide training of the highest standard to researchers using EMBL's scientific services. Close attention is given to providing a rapid response to user requests, offering free access or affordable prices where necessary, and the full integration of EMBL's services with the scientific objectives of EMBL and its collaborators.

EMBL's scientific services are considered a dynamic part of EMBL. Developing or introducing services is a continuous process shaped by the needs of various scientific communities. The research areas described in chapters 2–7, as well as EMBL's new technology developments, will spur the provision of new services in the future. Through significant investment in the next EMBL Programme, EMBL plans to strengthen and expand its scientific services (Figure SS0). For **experimental services**, including two major European synchrotrons, this will increase the capacity to host scientists visiting EMBL sites from member state institutions and beyond. Setting up the EMBL Mobile Services will also extend and deliver the benefits of EMBL's scientific services directly to the member states. For **data services**, which include over 40 data resources maintained by EMBL's European Bioinformatics Institute (EMBL-EBI), it is planned to deepen and further integrate these

information infrastructures, particularly those relevant to human and planetary research questions, as driven by research and user communities. A growing need, communicated by both the internal and external research community, is the **coordinated integration of scientific services**. External users will be able to leverage EMBL's internal collaborations and efficient workflows to answer their own research questions.

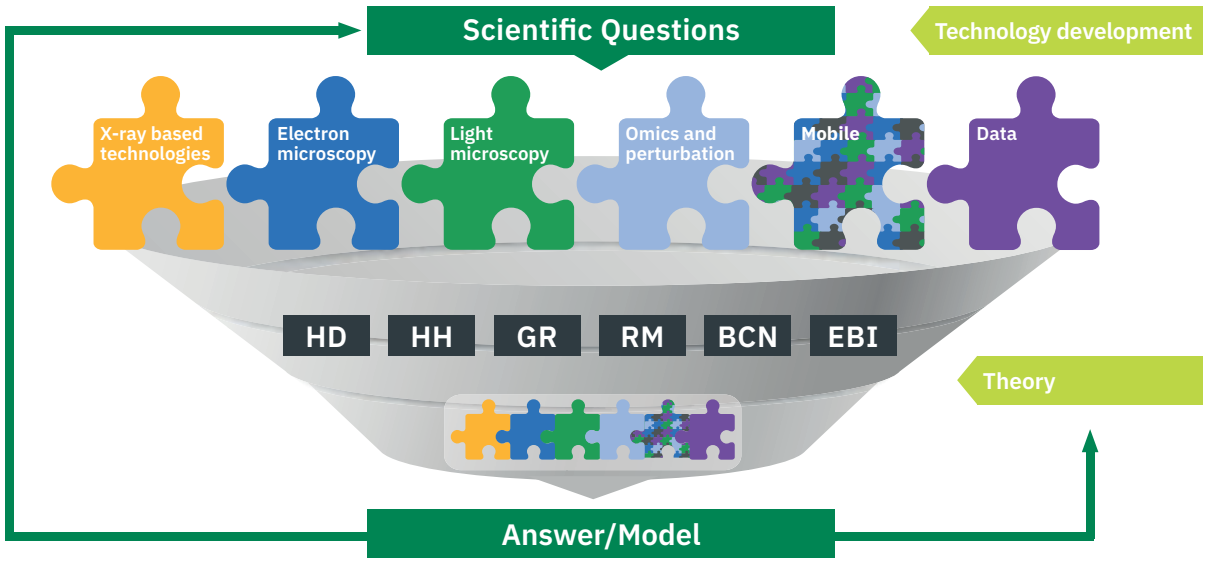


Figure SS0 | EMBL's scientific services.
EMBL's experimental and data services at EMBL Heidelberg (HD), EMBL Hamburg (HH), EMBL Grenoble (GR), EMBL Rome (RM), EMBL Barcelona (BCN), and EMBL-EBI (EBI) provide increasingly integrated, expert services to users in EMBL member states and around the world.

Structural Biology and Imaging Services

An understanding of life comes from knowing where and how molecules such as proteins, nucleic acids, and metabolites interact inside cells and organisms. Understanding such interactions requires structure–function studies of the individual components in the context of the larger molecular machines they form, and in the dynamic biological systems to which they belong.

EMBL's structural biology and imaging services encompass a range of methodologies, which utilise infrastructures across EMBL's sites. These services include X-ray beamline, cryo-electron microscopy (cryo-EM), and light microscopy services at EMBL Barcelona, EMBL Grenoble, EMBL Hamburg, EMBL Heidelberg, and EMBL Rome. Services like these enable scientists at EMBL and across its member states to answer research questions in multiple areas of molecular biology. Between 2015 and 2019, there was an annual average of around 4,300 user visits to EMBL's structural biology and imaging services. Collectively, these services cross scales, from atomic to optical resolutions (Figure SS1). However, the resolution boundaries between methods are becoming blurred as technologies continue to develop. New advances in X-ray imaging technologies make it possible to resolve millimetre-sized specimens to nanometre resolutions, cryo-EM allows proteins to be atomically represented within their cellular environment, and super-resolution microscopy now routinely breaks the light diffraction limit and enables the dynamics of protein complexes to be visualised in their natural habitat.

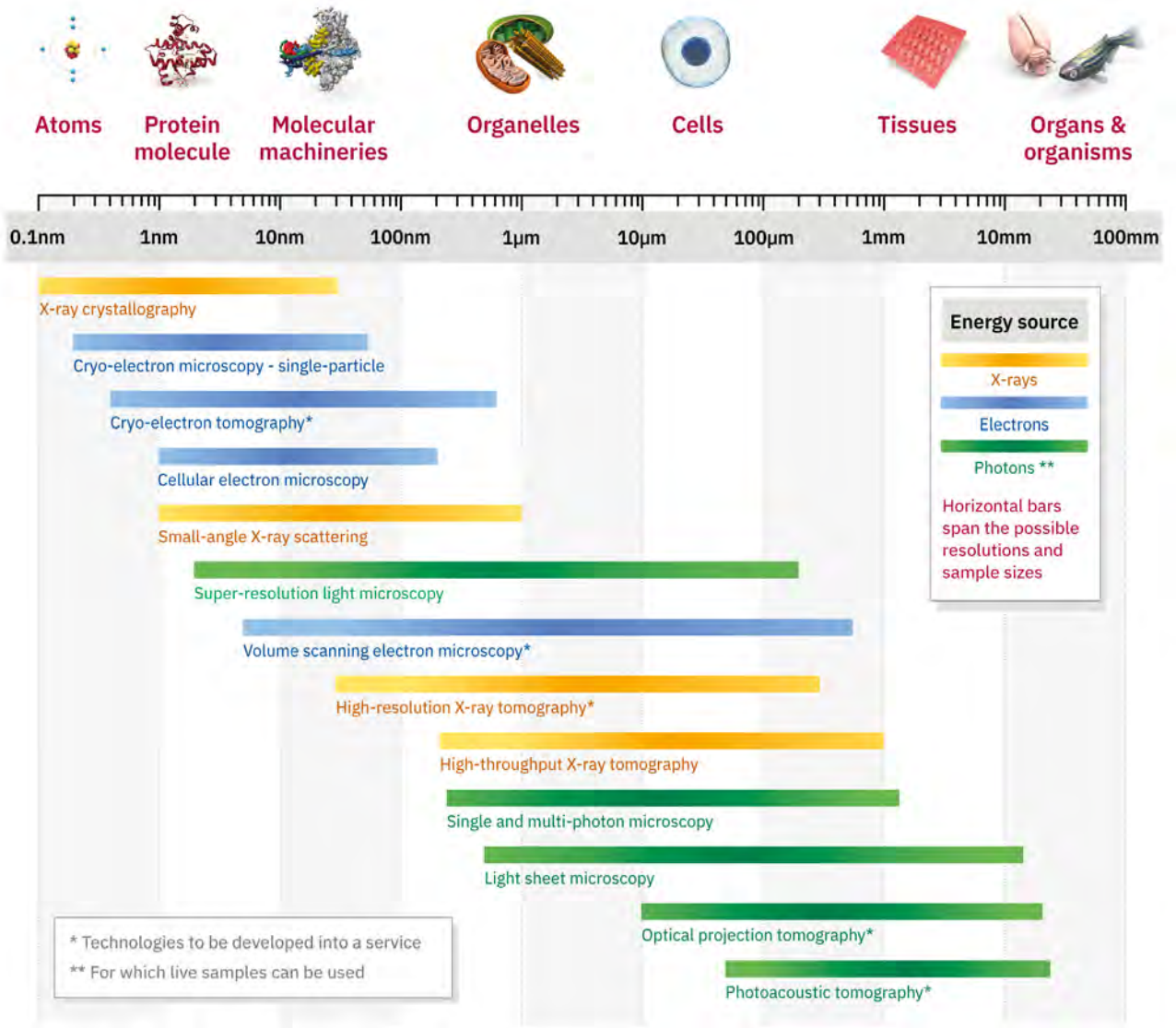


Figure SS1 | Structural biology and imaging across scales.
EMBL services in structural biology and imaging allow scientists to visualise biology across scales, from small living animals or entire organs, via cells and subcellular organelles, all the way down to biomolecules that can be resolved at the level of individual atoms. This offers the potential for scientists to acquire unique datasets, enabling the development of theoretical or mathematical concepts that can be used to understand and possibly predict the behaviour of biological systems, ideally in a way that integrates all information into one cross-scale model.

EMBL has led, and continues to lead, some of the key technological advancements in the fields of structural biology and imaging that have improved the resolution and penetration of technologies, understanding of molecular dynamics, and the ability to integrate different technologies. As part of EMBL's services, EMBL aims to make these cutting-edge, cross-scale technologies – which are usually embedded in the fields of mathematics, physics, and engineering – accessible to biologists who may only have foundational knowledge of these principles. By offering a range of technologies as services, EMBL can (1) advise biologists on the best technology to use; (2) assist in the preparation of samples; (3) support the optimum use of instruments and help with processing and interpretation of data; (4) help integrate and combine results obtained by different technologies; (5) support the deposition of data and models into publicly available data repositories. By fully supporting the use of complex experimental apparatus and by interfacing different scientific disciplines, EMBL services help to make it easier for biologists to use a range of structural biology and imaging techniques to answer complex biological research questions.

During the coming decade, EMBL infrastructures will leap forward at EMBL Heidelberg with the launch in 2021 of the EMBL Imaging Centre, which will implement high-end light and electron microscopy technologies, and at EMBL Grenoble and EMBL Hamburg with the upgrades of the synchrotrons they use. Fourth-generation synchrotron beams will be available to service users at the European Synchrotron Radiation Facility (ESRF) in Grenoble from 2020 onwards, and later in the decade at PETRA IV in Hamburg. The fourth-generation synchrotron beams will bring new opportunities for applications in life science research, due to substantial increases in beam brilliance, intensity, and coherence. The benefit of staggering these infrastructure upgrades is that EMBL can progressively develop and prototype technologies suitable for fourth-generation synchrotrons before launching highly refined technologies as services.

Structural Biology Services

Structural biology technologies allow biological molecules to be resolved at the level of individual atoms, making it possible to understand the biochemistry of life, how this biochemistry can be influenced in disease situations, or how it can be put to work in biotechnological applications. Such information is in high demand in the context of rational drug design and will be increasingly important in the development of enzymes optimised for industrial green chemistry and white biotechnology processes. While established structural biology technologies and services at EMBL Hamburg and EMBL Grenoble utilise the synchrotrons on the German Electron Synchrotron (DESY) campus and at the ESRF, respectively, the Imaging Centre at EMBL Heidelberg will further expand EMBL's cryo-EM service portfolio. By applying cryo-EM to single particles, it's possible to visualise large molecular assemblies at atomic resolution. Tomographic imaging with electrons also allows the visualisation of the cellular context in which macromolecules and their assemblies function. It therefore provides an important bridge between structural biology technologies that can resolve atomistic details of macromolecules and their interactions, with a wide range of imaging technologies that provide contextual information at various resolution levels. It is the combination and integration of structural biology and imaging technologies that allow us to zoom all the way from the organismal to the atomic level. Furthermore, high flux X-ray sources such as the new generation synchrotrons and the European X-ray Free-Electron Laser (XFEL) allow time-resolved studies of molecular dynamics and biochemical reactions.

EMBL is unique in its ability to offer a complete range of structural biology approaches and supporting infrastructures – in particular for preparing samples in a format suitable for a chosen technology – to support the external research community. Across EMBL there are expert data-producing sites that develop and use highly integrated approaches combining X-ray crystallography (MX), cryo-EM (single-particle, cryo-electron tomography), small-angle X-ray scattering (SAXS), mass spectrometry, and structural modelling and predictions to obtain meaningful biological insights. These approaches can be shared for complementary technologies such as the use of neutrons for probing biological materials as available in present and future research infrastructures including the Institute Laue-Langevin (ILL) and the European Spallation Source (ESS). EMBL-EBI as a data-hosting site maintains resources for data originating from a range of structural biology and imaging techniques. These services are embedded in European time-limited access programmes such as iNEXT, iNEXT-Discovery, and long-term ERIC infrastructures such as Instruct-ERIC, Euro-BioImaging, and ELIXIR (Chapter 13: Integrating European Life Sciences).

Future structural biology service plans will build on existing activities and further integrate complementary structural biology services across EMBL's sites. EMBL strives to develop, combine, and integrate the technologies necessary for structure determination in ways that can be made readily accessible at a user facility. Structure determination should be a **routine measurement for non-specialist users**, while specialist users should be enabled to **push methodologies to their limits** to extract the best possible data from difficult samples. One of the strengths of both EMBL Grenoble and EMBL Hamburg, as well as the forthcoming Imaging Centre and related core facilities, is the synergy between research, service, and instrument development

activities. This synergy allows for rapid user-driven development of new experimental strategies, with solutions frequently transferred into the commercial sector for the benefit of the worldwide user community.

During several decades of service provision in MX and SAXS, EMBL Hamburg has expanded its services, which began with beamline access, to a comprehensive package comprising sample preparation, data collection, and data evaluation. In fact, many of the software packages initially developed at EMBL Hamburg for dealing with MX or SAXS data are now in use by thousands of researchers and many facilities worldwide. In both MX and SAXS, the increases in sample throughput and data quality have recently made it possible to expand the types of measurement that can be made, from single measurements on a single sample to ensembles of related measurements on many samples. In this context, pump-probe time-resolved measurements of dynamic processes inside macromolecules have recently become feasible and are now being implemented as services (Figure SS2). Successful time-resolved experiments on a synchrotron beamline will also be an ideal preparation for executing similar experiments on X-ray free-electron lasers (XFELs). Upcoming improvements in beam quality, achievable both by deploying novel X-ray optics and by the upgrade of the synchrotron, will further improve throughput and – most importantly – the spatial and temporal resolution reachable with a given sample.

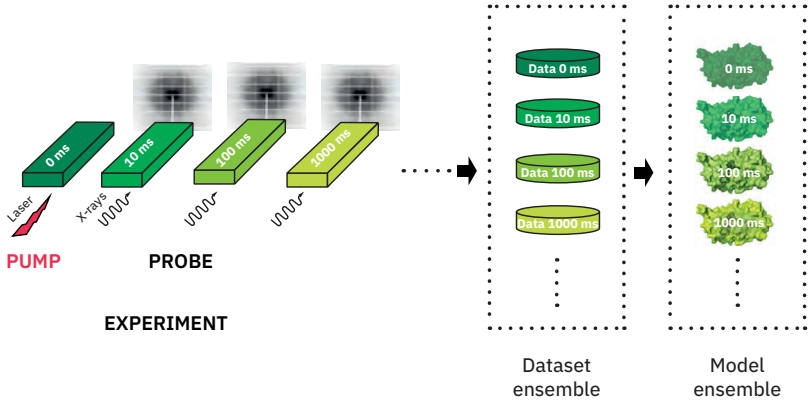


Figure SS2 | Time-resolved pump-probe serial crystallography.
Crystals containing protein molecules and ready-to-go substrate molecules (green) are 'pumped' with a laser (red) to start a chemical reaction inside the crystal. The crystal is then 'probed' by an X-ray pulse (black) and the corresponding diffraction pattern is recorded. By varying the delay between the pump and the probe stages (here shown at 10, 100, and 1000 ms), different time points of the reaction proceeding inside the crystal can be recorded. Complete 3D datasets for each time point, as assembled from extended serial exposures of many crystals, can then be combined into a movie of the 3D protein molecule in action. In contrast to traditional approaches, in which a single model depicts one specific static state of a molecule, a time-resolved series of measurements gives rise to dataset ensembles and model ensembles that correspond to the frames of a movie.

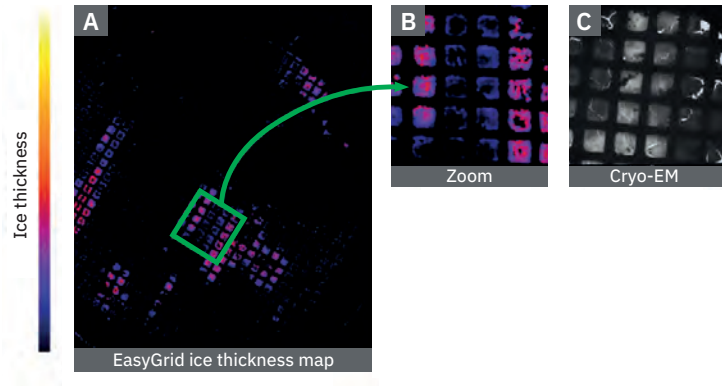
EMBL Grenoble continues to play a central role in the development of automated sample handling systems. For example, the development of technology to prepare and handle cryo-EM samples in an automated fashion is currently underway with the development of EasyGrid technology (Tech Dev Box TD1_SS). EMBL Grenoble has built expertise in this area by developing automated sample handling systems for both MX and SAXS, which are deployed on the seven structural biology beamlines at the ESRF in Grenoble and at PETRA III in Hamburg. In fact, the diffractometers developed at EMBL Grenoble have been successfully commercialised and are in use on most synchrotrons worldwide. The unique and innovative CrystalDirect technology for crystal handling and harvesting, developed at EMBL Grenoble, has closed the last gap in a fully automated pipeline to determine ensembles of crystal structures bound with different ligands or fragments and explore the chemical landscape of drug target molecules. The opening of the world's first fourth-generation synchrotron source at the ESRF in 2020, after the successful EBS upgrade, offers new challenges and opportunities for exploiting the ever-higher brilliance and more coherent beams. The fully automated beamline MASSIF is being upgraded for faster operation and with an online CrystalDirect device. The next generation of an automated and versatile SAXS sample changer will be deployed, with advanced

microfluidics and spectroscopy options. In addition, EMBL Grenoble is building an ultrafast and ultra-precise diffractometer for a new beamline that is under construction, which will enable studies of microsecond dynamics in crystals using serial crystallography methods.

During the past decade, improvements in hardware and software have led to a dramatic increase in the resolution that can be achieved by single-particle cryo-EM and cryo-electron tomography (cryo-ET). As a result, single-particle cryo-EM has rapidly become the method of choice for structure determination of individual proteins and protein complexes that are larger than 100 kDa. This is because, in many cases, a similar resolution to X-ray crystallography can be achieved without the need for growing well-diffracting crystals. These developments, sometimes referred to as the 'resolution revolution' in cryo-EM, have changed the landscape in structural biology and led to an increased demand for access to high-end cryo-EM instrumentation, and equally for additional training in cryo-EM. EMBL has reacted promptly to these increased demands. Since 2017, EMBL Heidelberg has operated its Cryo-EM Service Platform, which provides external users with access to high-end cryo-EM instrumentation in preparation for the EMBL Imaging Centre, which will open in the summer of 2021. EMBL Grenoble participates in a state-of-the-art cryo-EM facility run jointly by the institutes of Grenoble's EPN science campus. The facility operates like a protein crystallography beamline for projects requiring high-end cryo-EM data collection, with free access provided for projects that are accepted following peer review. It is likely that this facility will acquire additional microscopes in the future. At EMBL Hamburg, access to cryo-EM instrumentation for in-house research is currently provided through the Centre for Structural Systems Biology (CSSB), with discussions ongoing to create extra capacity for external users.

Technology Development Box TD1_SS | EasyGrid.

Through the project EasyGrid, the Cipriani Team at EMBL Grenoble is developing ways to automate the process of preparing cryo-EM sample grids. Automation will improve reproducibility of prepared samples, thus facilitating their optimisation to collect the best possible data. EasyGrid targets low sample volumes, multiple samples on the same grid, time-resolved vitrification and sample quality check after vitrification. Adequate sample ice thickness is a mandatory prerequisite to collect data, hence an ice thickness map with 10 nm resolution **(A)** will be provided with each prepared grid. An EasyGrid ice thickness map allows the visual correlation between a zoomed-in region of the EasyGrid ice thickness map **(B)** and the corresponding cryo-EM grid region image **(C)**, using a grid prepared with a Vitrobot. The EasyGrid nanometer-scale ice thickness maps can be used just after sample preparation to assess the quality of the grids, therefore making it possible to immediately reject poor grids, and efficiently process those with exploitable sample areas by directly positioning them in the electron beam. This will considerably reduce the time spent on screening grids at cryo-EM platforms and the time spent on setting up data-collections. Therefore, this development will not only facilitate the work of expert microscopists but will widen access of the cryo-EM imaging method by proposing an easy-to-use and reliable sample grid preparation and characterisation system.



Finally, the data flow between data producing, processing, and deposition sites will be improved as the size of structural biology datasets continues to increase, amounting to several terabytes a day, with levels of complexity ranging from single structures to ensembles. This necessitates the development of data-handling policies to cover data and metadata collection, raw data storage, data processing, and archiving (Chapter 8: Data Sciences).

EMBL-EBI is the European hub for the deposition and distribution of structural biology data from X-ray diffraction (PDBe) and electron microscopy (EMDB, EMPIAR). In addition, EMBL Hamburg hosts the Small Angle Scattering Biological Data Bank (SASBDB). EMBL therefore has the expertise and the mandate to drive innovation in data flows and data storage in structural biology. In addition to a strong background in technology development, EMBL's unique combination of research and service activities provides an ideal environment in which it's possible to address scientific aspects of the process of going from data to information to knowledge. One example of this is the Protein Data Bank in Europe (PDBe) at EMBL-EBI, which actively works with its worldwide PDB partners and with SASBDB on data validation issues, to improve the quality of deposited data, including data from integrative or hybrid methods. A second example is the PDBe Knowledge Base (PDBe-KB) which collates value-added annotations pertaining to PDB structures, including annotations on disorder, and is continuing to develop intuitive visualisation tools with specialised communities involved in shaping and testing them.

Imaging Services

The aim of imaging services at EMBL is to provide access to state-of-the-art technologies and associated operational expertise. All EMBL imaging services excel in providing access to the latest cutting-edge light and electron microscopy instruments and developing new imaging technologies and services in close collaboration with leading industry and instrument developers, either at EMBL or in its member states. Importantly, users are trained to expertly and independently prepare and image their samples, before processing and analysing their data using standardised and quality-driven methods. These key points enable EMBL and EMBL member state scientists to stay at the competitive edge of the rapidly developing imaging field. In doing so, many more pressing and exciting research questions can be answered.

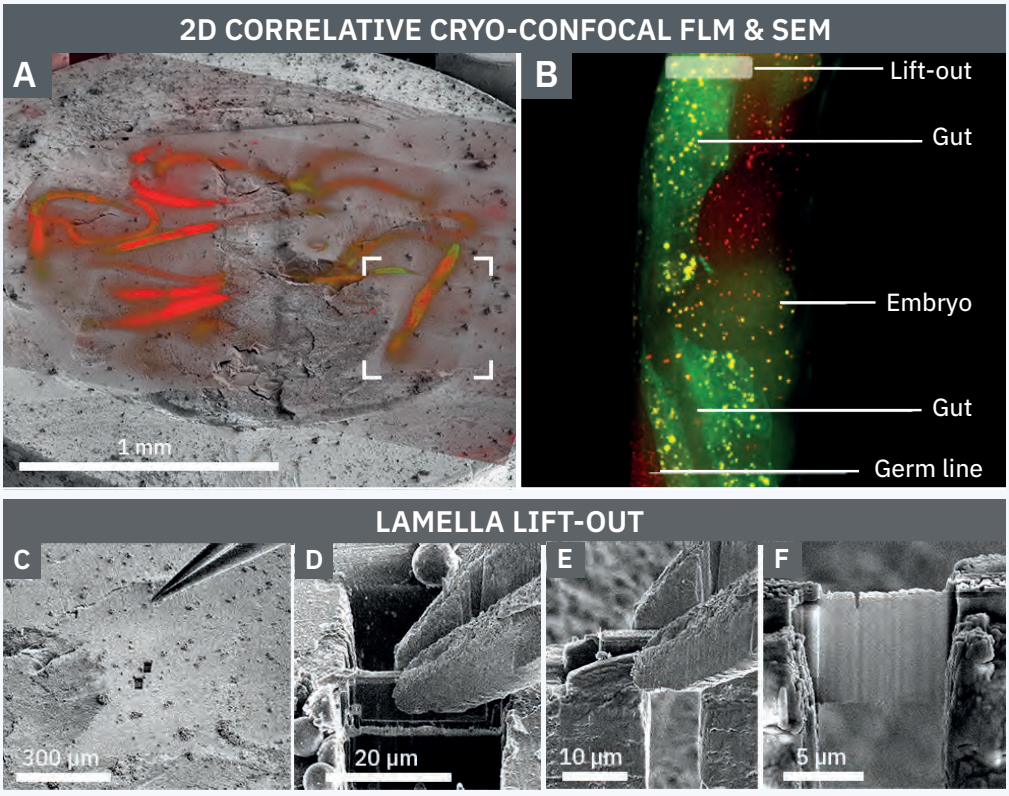
As a result of this approach, EMBL's imaging services offer a wide range of highly advanced techniques that cover multiple dimensions and scales (Figure SS1). **Ultrastructure** can be assessed at high throughput by automated transmission electron microscopy (TEM) and linked to subcellular localisation of proteins by correlative light and electron microscopy (CLEM). **Cellular functions** can be studied, for example, by automated cross-correlation spectroscopy to quantify the dynamics of molecular interactions, as well as high-throughput confocal and super-resolution fluorescence microscopy to allow highly multiplexed imaging-based experiments at large scale, up to the systems level. **Mesoscopic samples**, such as a developing limb bud or organoids, which cross the micro- and macroscopic scales, can also be imaged using EMBL's services. Two technologies integral to this field of imaging are optical projection tomography (OPT) and selective plane illumination microscopy (SPIM).

EMBL's imaging services will further develop during the next EMBL Programme. On the one hand, imaging services will increase their capacity, principally to follow the development of their user base, triggered largely but not exclusively by the newly launched Euro-BioImaging ERIC, as well as by the implementation of new mobile services (see below). EMBL will also strive to integrate new technologies that are reaching a sufficient degree of maturity to be offered as imaging services. These services, detailed below, further strengthen the uniqueness of EMBL's imaging services, which cover a wide range of dimensions and scales using the latest multimodal technologies (Figure SS1).

Cryo-ET and **cryo-CLEM** are techniques at the interface between structural biology and imaging services. With these modalities (Tech Dev Box TD2_SS), the structure of molecular assemblies is accessible *in vivo*, in integrated systems such as cells, tissues, or small model organisms. Currently in an advanced development phase that involves EMBL research groups and industrial partners, these techniques are soon expected to mature in terms of robustness and throughput, allowing them to be included as part of EMBL's imaging services in the near future.

Technology Development Box TD2_SS | Multimodal cryo-microscopies.

Integrating atomic resolution uniquely obtainable by cryo-electron tomography (ET) with molecular cell biology to advance the new field of structural cell biology requires the development of correlative and multimodal microscopy technologies. EMBL scientists and services are streamlining technological workflows to enable and broaden access to cryo-ET on complex biological systems. In the figure below, light and electron microscopies performed at liquid nitrogen temperature are used to target and identify cells, organelles or macromolecules inside frozen-hydrated whole model organisms **(A and B)**. Such maps can be utilised to navigate the sample and micromachine it by means of a cryo-focused ion beam (FIB) precisely at the site of interest in order to create and lift a lamella out **(C and D)** ready for cryo-ET where the final lamella is thinned to 200 nm **(E and F)**. Another complementary approach currently being developed is to image the sample surface while sputtering away material by cryo-FIB/SEM tomography (not pictured). Such a slice-and-view approach reveals the anatomy of the specimen while progressing within its core and is expected to enable a precise positioning of a lamella and provide contextual information on the biological specimen at a continuous spectrum of resolutions. It is envisioned that developing the interface between these complementary multimodal imaging approaches within a single instrument will lend an enormous discovery potential to any life, biomedical and material sciences study worldwide.



Volume scanning electron microscopy (SEM) techniques are also becoming more accessible, due to hardware improvements in terms of image acquisition, speed, sensitivity, and resolution, and through newly developed workflows adapted to an increasing number of fundamental research topics in the life sciences. Importantly, the development of **3D CLEM** strategies will integrate volume SEM with other cutting-edge imaging modalities such as light-sheet and mesoscopic imaging or 3D X-ray tomography, adding to the portfolio of techniques that contribute to an understanding of the links between structure and function.

3D super-resolution light microscopy has developed in recent years to a remarkable degree, and pioneering work in the field has made it possible for resolutions of a few nanometres to be achieved. Examples are 4Pi single-molecule localisation super-resolution microscopy or MINFLUX technology. These technologies further close the gap between light and electron microscopy and may in the future offer the unprecedented opportunity to observe structural changes of molecular machines at work in their natural habitat. EMBL will further develop these technologies in collaboration with leaders in the field, such that they mature sufficiently to be offered as services to the life sciences community in the forthcoming EMBL Imaging Centre.

New technologies for imaging mesoscopic samples, such as **photoacoustic tomography** complemented with advanced **multiphoton fluorescence microscopy for deep tissue penetration**, are being developed by EMBL's microscopy technology development teams. In the future, these technologies are expected to become sufficiently robust to further complement the EMBL imaging service portfolio in the mesoscopic range.

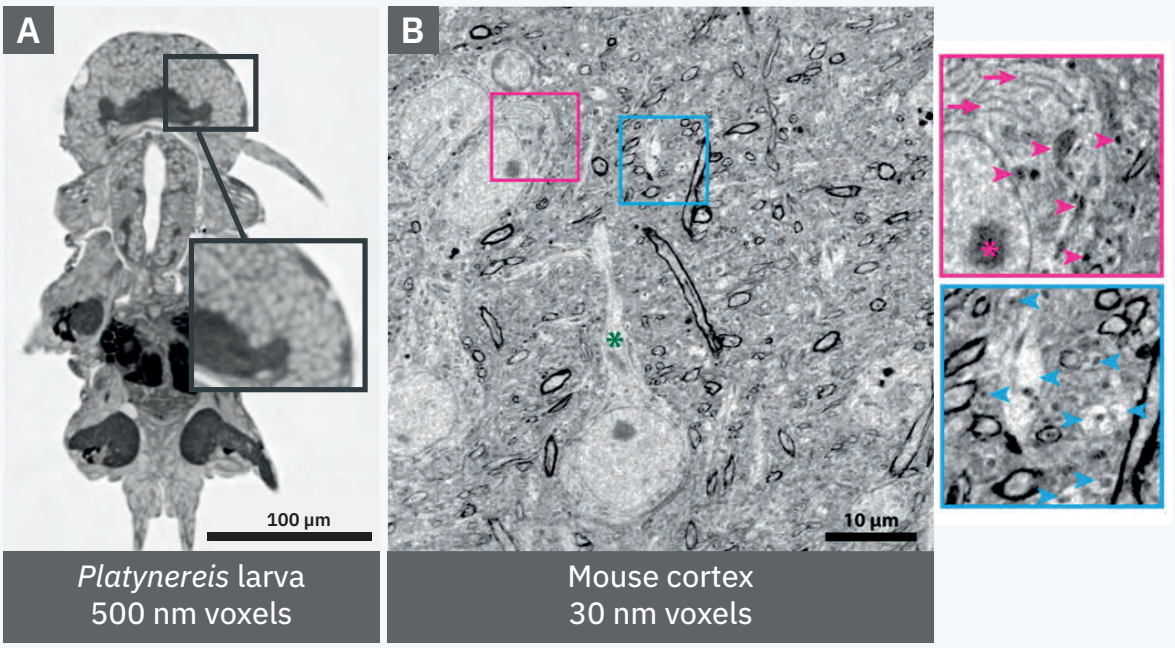
In **3D synchrotron X-ray tomography**, groundbreaking experiments have demonstrated the use of high-energy (> 6 keV) X-rays for imaging large biological samples without sectioning, from organisms all the way to individual neuronal processes, reaching 30 nm voxel size in the best cases. High-energy X-rays offer deep penetration of materials, so it is possible to apply tomographic methods. These provide 3D images, and in favourable cases enable time-resolved studies *in vivo*, though at a lower resolution. Fourth-generation synchrotrons will offer high-energy X-rays with unprecedented quality in terms of beam intensity, homogeneity, and coherence. Taken together, this X-ray 'resolution revolution' will enable a leap in our capacity to uncover ultrastructural features of large intact biological samples rapidly and systematically (Tech Dev Box TD3_SS). EMBL is pursuing several avenues to improve X-ray tomography sample preparation and automation, and to apply it to neurobiology.

At EMBL Hamburg, X-rays from PETRA III have been used to image small organisms in their entirety at cellular resolution in a high-throughput regime, paving the way for large-scale studies of morphology and morphological variation. Full-field X-ray tomography can be pushed to a resolution where neuronal connectivity can be imaged with large fields of view. In nanoprobe mode, X-rays can be used to extract information on the spatial distribution of specific chemical elements inside cells.

Successful X-ray tomography of biological samples demands new developments to create robust sample preparation methods and strategies and technical implementations for processing and analysis of terabytes of raw data per study. With its presence at two world-leading synchrotron facilities, EMBL is participating in the ongoing upgrade of the ESRF and the forthcoming upgrade of PETRA IV, and is in a unique position to exploit the opportunities offered to the life sciences by X-ray imaging.

Technology Development Box TD3_SS | The 'resolution revolution' in X-ray tomography.

Electron microscopy (EM) remains the gold standard for obtaining nanometre-resolution ultrastructure of biological tissues. However, EM requires laborious tissue sectioning and assembly, which limits its use to small sample numbers and sizes and anecdotal experiments. Recent work at the ESRF has shown that high-energy X-rays and cryo-tomography techniques can be harnessed to obtain 3D images at nanometre resolution from large, millimetre-sized tissue samples in several hours instead of the months required for imaging similar samples by EM (Pacureanu *et al. bioRxiv* 2019). This resolution breakthrough represents a unique opportunity for EMBL to use its expertise in beamline sample preparation and automation and high-content image analysis to apply this technology in the life sciences. In particular, X-ray holographic nanotomography has great promise for mapping neural circuit ultrastructure and connectivity in the brain, and EMBL Rome has embarked on a series of collaborations with the Pacureanu lab at the ESRF to develop sample preparation, image analysis, and molecular genetic labelling and post-imaging mRNA quantification approaches for neurobiology applications. EMBL aims to integrate these technologies into a 3D tissue ultrastructure imaging service available to the wider neuroscience community and for further applications in clinical tissue diagnosis.



Rapid tissue ultrastructure imaging without sectioning by X-ray tomography. **(A)** Section through a medium-resolution (500 nm voxels) 3D X-ray tomogram of the marine worm *Platynereis dumerilii*, based on data collected on EMBL's P14 beamline at PETRA III. Inset shows resolution of individual neurons in the worm brain. **(B)** Section through a high-resolution (30 nm voxels) 3D X-ray tomogram of mouse cortex based on data collected on ESRF's ID16A beamline. Insets show neuronal processes and organelles imaged at near-EM resolution (Pacureanu *et al. bioRxiv* 2019).

EMBL Imaging Centre

The EMBL Imaging Centre, planned to open for service provision in the summer of 2021, represents a new EMBL service unit for the most advanced imaging technologies integrated across scales (Figure SS3). The EMBL Imaging Centre will cover methods from the highest-resolution cryo-EM to the latest intravital light microscopy technologies, including academically developed technologies and workflows not yet commercially available. The main mission of the EMBL Imaging Centre is to make the most advanced microscopy technologies available as quickly as possible to a broad scientific international user community from both academia and industry.

Figure SS3 | Rendering of the EMBL Imaging Centre.

A powerful and synergistic portfolio of imaging technologies will be available for external service users (Figure SS1). To integrate technologies, the EMBL Imaging Centre will have a special emphasis on combined, or 'correlative', imaging that crosses the scales of biology. This includes electron microscopy technologies covering all scales, from single-particle cryo-EM and cryo-electron tomography, to (cryo-)CLEM, cellular or tissue volume electron microscopy (FIB-SEM, SBEM). The EMBL Imaging Centre will specialise in light microscopy technologies covering all scales from the latest (cryo-)super-resolution technologies to high-speed and high-resolution live-cell, intravital, and deep tissue imaging, with a focus on non-commercial, academically developed technologies. Users will receive expert, tailored project support from sample preparation to image data analysis, building on the extensive experience gained in EMBL's current microscopy core facilities. The EMBL Imaging Centre will provide access to new technologies, as soon as they are robust, for up to 300 users per year. The services offered will respond dynamically to the needs and expectations of visitors.

A fundamental aim for the EMBL Imaging Centre is constant improvement and development of new imaging and correlative technologies and workflows. It is envisioned that the EMBL Imaging Centre will become a European hub for introducing the latest pre-commercial imaging technologies to groundbreaking life science applications. This will be achieved jointly by the EMBL Imaging Centre's service staff, its industry partners, EMBL research groups aiming to develop groundbreaking new imaging technology, and the European microscopy technology developer community. The EMBL Imaging Centre will invite technology developers and their teams from EMBL member states for extended visits to EMBL, with the aim of testing and improving their latest technology developments alongside cutting-edge applications devised by EMBL scientists or external service users. Such visits could ultimately lead to the establishment of these technologies as services for the life sciences community at very early stages of their development. These visits will also help external technology developers to evaluate and improve their technologies for applications in the life sciences, and will allow dissemination of the new and improved technologies, possibly via spin-out companies or collaborations

with industrial partners (Chapter 12: Innovation and Translation). The priorities for these new developments will be driven by biological applications; developments will range from novel sample preparation (Tech Dev Box TD2_SS), via single instrument development or workflow development to correlate or integrate several instruments, to novel image analysis tools.

A New Level of Image Data Analysis Services

EMBL scientists and, increasingly, the external users of EMBL imaging facilities are producing rapidly increasing quantities of high-quality bioimage data. In addition, the large-scale provision of services to produce bioimage data within the new EMBL Imaging Centre will create immediate needs for image data analysis in the user community from member states. Given both the size and complexity of the data produced by these cutting-edge imaging technologies, specialised analysis expertise and adequate tools are required to develop user-tailored solutions for the data analysis challenges occurring in user projects.

To address these demands and complement the data-generating imaging services, a new Bioimage Data Analysis Service Team will be established to align and extend the expertise of the excellent support staff already present at EMBL. This service team for bioimage data analysis will closely interact with the EMBL groups that develop new analysis methods, to provide these tools in a user-friendly manner to the internal and external user community. In the longer term, the tools and methods will be developed into robust services that can be hosted by EMBL as public services accessible via the cloud.

Structural Biology and Imaging Centre Access Platform

EMBL envisions a uniform, cross-site accession platform for EMBL structural biology and imaging services. This platform would enable easy user access and further increase EMBL's visibility as a service provider. The inclusion of the new EMBL Imaging Centre within this vision provides the opportunity to further integrate structural and imaging services.

The online platform will offer streamlined information and effective access to the range of technologies available at EMBL, spanning scales from atomic-resolution techniques such as cryo-EM to high-resolution light microscopy up to the mesoscopic scale. While the platform would offer a single point of access to services, the entry and exit points for specific service pipelines could be chosen by the user. Experienced users could limit their use of services to data collection only, while less experienced users could start with preparation of samples and be supported all the way through processing and analysis of the data. Underlying the public-facing portal would be a unified review and instrument access procedure. This would include project evaluation schemes to review a project's relevance, feasibility, and safety before enabling flexible and rapid use of instruments at EMBL sites. Where possible, joint user offices, management tools, and common managerial structures would be put in place. Importantly, joint activities would include concerted actions to secure funding for user programmes across all sites; one example of this is the involvement of all sites in iNEXT-Discovery. These procedures aim to harmonise activity across EMBL sites while taking account of the specific features of local infrastructures.

Multi-omics Services

The multi-omics services at EMBL include all techniques that relate to the study of genes and their functions, namely DNA sequencing, transcriptomics, proteomics, and metabolomics. Multi-omics services offer state-of-the-art technologies and expertise for cutting-edge functional genomics analyses, taking advantage of the continued revolutionary developments in next-generation sequencing, mass spectrometry, and genome editing and are integrated across EMBL's core facilities. As with all of EMBL's scientific services, EMBL's core facilities ensure that users gain access to **infrastructure, invaluable expertise, and quality assurance**. These three key principles ensure that EMBL's core facilities support and advance research and technology development at EMBL and within its member states.

The Genomics Core Facility prepares and sequences DNA and analyses the resulting genomic data, to allow the referencing, identification, and further analysis of biological samples from organisms ranging from bacteria to humans. The facility uses state-of-the-art technologies to provide services in gene expression analysis (RNA-seq), upstream regulatory analysis (sequencing samples of ChIP-seq, DNase I, ATAC-seq, DNA methylation), and DNA whole-genome sequencing. Robust instrumentation infrastructure allows the preparation of libraries for various applications.

The Proteomics Core Facility provides the infrastructure needed to identify and characterise proteins, namely mass spectrometry for liquid chromatography and mass spectrometry experiments. This is complemented by systems for chromatography and electrophoresis, for protein and peptide separation. Regular upgrades to instrumentation allow deeper and faster proteome analyses, increasing the capacity of the facility and providing opportunities to develop new services.

The Metabolomics Core Facility provides services for the analysis of metabolites and lipids. The facility is equipped with high-resolution mass spectrometry systems coupled with liquid chromatography and technologies for the separation and detection of various classes of small molecules, metabolites, and lipids.

Advances in Single-cell Genomics Services

For decades, it has been standard practice to measure molecular signatures, such as changes in response to environmental stimuli or drug treatment, in samples that contain thousands or even millions of cells. This results in average measurements, reflecting the general response within the entire cell population. All information about how individual cells respond is lost, and heterogeneity within the population is masked. Single-cell genomics technology, which has matured considerably in recent years, solves these issues.

Single-cell genomics is providing the first genome-wide views of cell fate trajectories during embryogenesis, enabling the discovery of new cell types, revealing the extent of tumour heterogeneity and diversity in response to drug treatment, and uncovering how gene expression responses cluster across cells. The field is progressing rapidly, moving beyond single-cell RNA-seq to measuring upstream parameters in the regulation of gene expression as well as multiple parameters from the same single cell (often referred to as single-cell multi-omics). A number of groups at EMBL are also developing new experimental methods and new computational models and tools to analyse molecular signatures at the single-cell level. This includes genetic (DNA), epigenetic (chromatin features), and gene expression (RNA) information from the same single cell.

EMBL's multi-omics services, in collaboration with the Flow Cytometry Core Facility at EMBL Heidelberg, support users by providing expertise and services in scRNA-seq and scATAC-seq using commercial platforms such as 10x Genomics, as well as sequencing services and library preparation help for advanced users

developing new single-cell methods (eg. scATAC-seq, Strand-seq). In addition to providing access to these methods to scientists from EMBL member states, the facility organises hands-on training courses and has formed a network of core facility heads throughout Europe to facilitate open exchange and best practice.

EMBL is a key driver within the single-cell genomics field, through innovative research and by leading and participating in local and international community initiatives such as the Single Cell Center Heidelberg and the Human Cell Atlas (HCA), respectively. The HCA collaboration has helped highlight that the many diverse cell types in tissues and organisms, including newly discovered cell types, can be described as belonging to a continuum of cell states.

Emerging Developments in Spatial Omics

Although single-cell measurements are a powerful tool in helping scientists understand the granular responses of a sample, they typically involve dissociating cells within a complex tissue or sample, and thereby lose all spatial information on where the cell resided in its natural environment. Spatially resolving genomics at single-cell resolution can address longstanding questions in the fields of developmental biology, infectious biology, neurobiology, and human biology and disease. For example, spatial genomics can provide an understanding of how a cell responds to internal and external stimuli as a function of its position relative to other cell types or the source of a signal. This can be carried out in the context of an embryo's development, host–pathogen interactions, or studying heterogeneity in responses to drug treatment within a tumour. Moreover, single-cell spatial omics offers the possibility of linking gene expression of individual cells to phenotypic data derived from other tissue-imaging approaches.

The need to answer questions in areas such as these drives demand by researchers, both within and outside EMBL, pushing EMBL to integrate imaging techniques with various omics technologies. Once a proof of concept has been demonstrated, these spatial transcriptomics methods will be developed further into a semi-automated, robust, integrative service. The Metabolomics Core Facility already provides expertise in the emerging field of spatial metabolomics, based on high-resolution imaging-based mass spectrometry. In the future, EMBL scientists will develop spatial multi-omics methods to integrate spatial measurements with many parameters from the same single cell *in situ*, for example by linking data on RNA, proteins, and metabolites with phenotypic information as it is obtained, for example by advanced functional tissue imaging methods.

Currently the Genome Biology Unit and core facilities are setting up a Genomics Technology Development team (GenTechDev) at EMBL Heidelberg to coordinate and foster activities across EMBL, integrating imaging and genomics methods to develop, optimise, and implement new single-cell spatial genomics technologies. As these technologies are very complex, requiring advanced expertise in single-cell genomics, state-of-the-art imaging, and high-level computing, the team will serve as a hub to integrate current efforts throughout EMBL. Technological advances being developed at EMBL include image-based sorting of samples (Figure SS4). This involves the automated, multiplexed 3D imaging of samples and novel data analyses which correlate and integrate quantitative microscopy data – including high-resolution X-ray imaging – and multi-omics data. It is expected that technology developed in the GenTechDev team will be used and further developed across EMBL sites in cutting-edge applications, and should lead ultimately to a level of robustness and demand such that EMBL's core facilities will offer the technology as a service to scientists at EMBL and users across EMBL's member states.

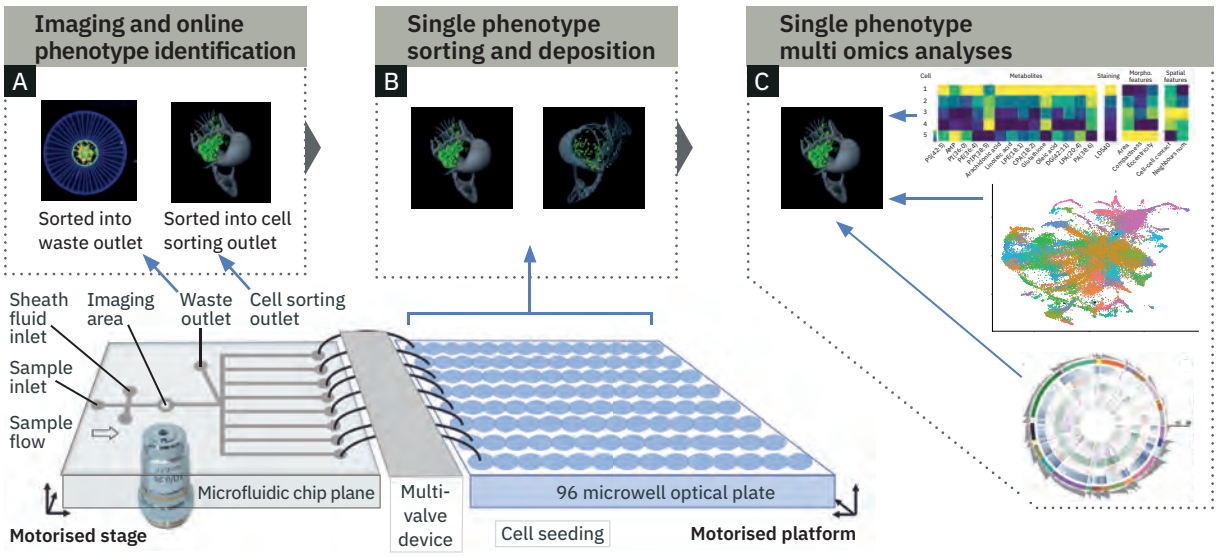


Figure SS4 | Imaging-based sorting and multi-omics analysis of single-cell phenotypes.
(A) Phenotypes of single cells are identified by automated high-resolution imaging and online image analysis. **(B)** Phenotypes are subsequently sorted by microfluidics and deposited into multiwell chambers, **(C)** followed by single-cell omics analysis.

In Vivo Gene Editing Service

The exponential increase in scientists' ability to measure genetic variation has dramatically increased the need for new technologies to experimentally modify or edit genomes to understand their function. The advent of CRISPR-based genome editing tools is filling this gap, and EMBL has been at the forefront of applying these tools for cellular and organismal genome editing. It is now possible to make point mutations in DNA and RNA virtually at will, and to do this in living organisms in a cell-type-specific and region-specific manner using viral-based delivery methods. Moreover, CRISPR-based targeting modules can be used to bring enzymes or other molecular machines to specific genome locations as needed to modulate local genome function – a technique referred to as epigenome editing. EMBL aims to continue its leadership role in this area and to invest in service activities to ensure this expertise is available to member state laboratories where there may not be the infrastructure or experience needed to implement them. EMBL is also using these techniques to carry out a series of systematic genome editing and functional analysis projects that will make animal and cell line resources available. These resources promise to have a wider impact in the field of genetic and epigenetic gene expression control. While genome editing technologies are being developed and applied across EMBL, the Gene Editing and Embryology Facility (GEEF) at EMBL Rome will assume a central role as a service provider for innovative *in vivo* gene editing in mice.

Studying Human Genetic Variation in Mice

The GEEF has developed major in-house capacity to generate genetically modified mice using CRISPR-based tools. Targeted germline knock-in of point mutations, as well as protein tags and recombinases, is routine and is being offered to both internal and external researchers. For example, the GEEF is now regularly producing mice in which limited sections of the genome are converted into its human orthologue – so-called humanised mice – and then engineered to carry a human disease-causing genetic variant. Humanised mice are powerful tools to test the impact of human genetic variation on whole-animal physiology, and will be a critical part of the experimental toolbox for testing the causality of coding and non-coding genetic variants emerging from human genome-wide association studies (GWAS) and rare-variant association (RVAS) studies. They will

also be essential platforms for testing CRISPR-based therapeutic vectors (viral-based gene therapy) for the correction of disease-causing variants. Importantly, the GEEF offers a full service pipeline covering custom variant design, founder production, and animal validation.

Viral-based *In Vivo* Editing

Gene editing of adult somatic tissues is increasingly relying on viral-based delivery of CRISPR gene editing components. EMBL Rome's Genetic and Viral Engineering Facility (GAVEF) has developed high-titre adeno-associated virus (AAV) vectors for high-efficiency cell-type-specific CRISPR-based gene knockout in mice and tissue organoid systems. In addition, the facility offers a wide variety of lentivirus vectors for genome and epigenome editing in cell lines. GAVEF offers a full custom viral vector design and production service to internal and external researchers, which will be expanded as part of the new EMBL Programme.

Facilitating *In Vivo* Functional Genomics – Epigenetic Zoo

In a large systematic genome engineering project, EMBL researchers have produced a reference collection of mouse strains carrying catalytic point mutations in epigenetic readers, writers, and erasers, known as the epigenetic zoo. Until now, epigeneticists were largely limited to correlative studies linking epigenetic marks and gene expression changes. These mouse lines will allow for a systematic causal understanding of epigenetic marks. Importantly, viral gene knockout tools (see above) can be used to uncover epigenetic zoo catalytic-dead mutations in a tissue- and cell-type-specific manner to study epigenetic causality across cell types (Figure SS5). Mice and cell lines derived from the epigenetic zoo, as well as viral engineering tools, will be made available to the research community.

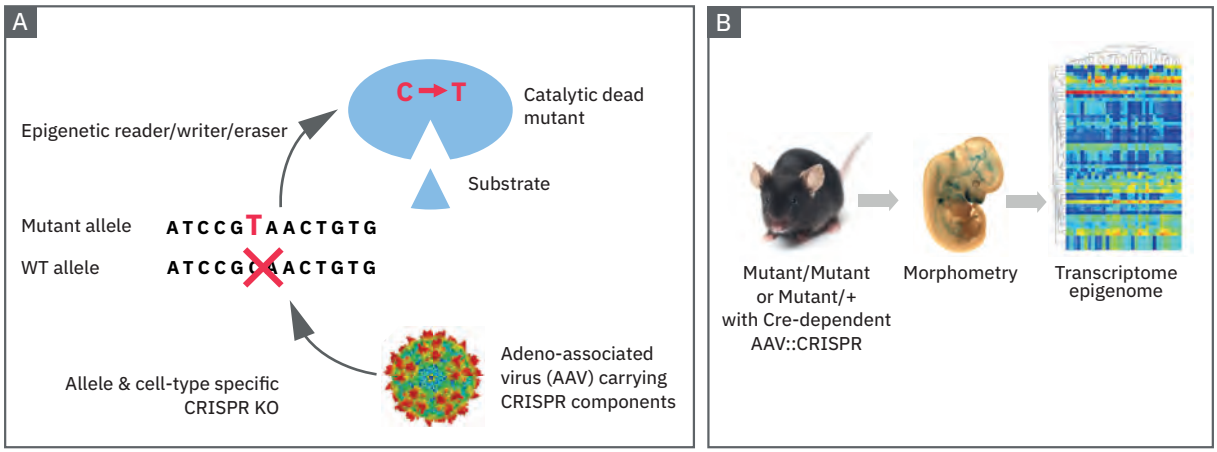


Figure SS5 | Establishment and characterisation of an epigenetic zoo.
(A) EMBL researchers have established a series of mouse lines carrying catalytic point mutations in epigenetic readers, writers, and erasers. Because these factors are an integral part of cellular chromatin, knockout mutations cause severe disruptions and catalytic point mutations are needed to uncover the functional role of the epigenetic mark they control. While homozygous mutant lines will be used to understand systemic functional deficits, heterozygous lines will be used to induce cell-type-specific disruptions of epigenetic marks by Cre-dependent viral-dependent CRISPR knockout of the wild-type allele. **(B)** Lines will be systematically characterised for morphological deficits using automated image quantification and for changes in the cell-type-specific transcriptome and epigenome.

Chemical Biology Services

The Chemical Biology Core Facility (CBCF) offers infrastructure and expertise for quantitative assay development, small-molecule screening, and support in medicinal chemistry to optimise compounds and give insights into protein function. When coupled with other facilities across EMBL, these services can form unique, high-throughput, and systematic workflows to identify, design, develop, and validate high-quality molecular probes to elucidate the molecular mechanisms of biological processes. Several research questions relevant to the new EMBL Programme can be addressed using EMBL technologies such as EMBL Grenoble's high-throughput crystallisation platform CrystalDirect, which enables screening of thousands of crystals for bound small molecule ligands from fragment libraries, or expertise from the Protein Expression and Purification Core Facility in the major recombinant protein expression systems. Capabilities to operate cross-site facilities will be further developed to provide screening campaigns and integrated experimental workflows for researchers at EMBL, in the EMBL member states, and beyond.

New **screening campaigns** directly related to the research areas outlined within the Microbial Ecosystems, Infection Biology, Human Ecosystems, and Planetary Biology chapters could enable scientists to characterise chemical libraries of pesticides, antibiotics, natural compounds, and steroids, as well as primary and secondary metabolites in the context of the health of environmental and microbial systems. Specific compound libraries such as these do not currently exist but would be of crucial importance for mass spectrometry-based analyses of field samples, and would be powerful references for scientists at EMBL and in EMBL member states. Undertaking these screening campaigns would allow EMBL and its service users to assess the effects, including toxicity, of compounds and their degradation products on communities and organisms under controlled environmental conditions (Chapter 7: Planetary Biology) and using **integrated experimental workflows**. This could provide a mechanistic understanding of the effect of environmental pollutants, agrochemicals such as pesticides or fungicides, natural compounds in different sources of food, or the effect of a number of food additives on ecosystems. It could also be used to identify tool compounds for drug discovery.

In human-related systems such as the gut microbiome, it could be possible to further determine the impact of these chemicals on their natural counterparts. For example, compounds that perturb the microbiome could be identified and their potential to alter the composition of the microbiome to improve human or animal health could be investigated. Initial proof-of-principle experiments, which show that the microbiome composition can be selectively modulated by prodrugs, have been published (Chapter 4: Microbial Ecosystems). The next step could be to synthesise **libraries of positive and negative regulators** of bacterial growth and test them for specificity. The ultimate aim would be to find methods to reduce populations of pathogenic bacteria in the gut, either through direct inhibition or indirectly by stimulating the growth of beneficial species. Metagenomics databases such as MGnify at EMBL-EBI can also be used as a starting point to identify organisms and enzymes with the potential to degrade long-lived organic environmental contaminants.

Another need for the CBCF is to support the **development of fluorescent molecular sensors**, for example by screening libraries of these sensors to optimise their performance. This would enable the most efficient and effective sensors to be developed and used in imaging-based experiments. Imaging technologies will remain fundamental for investigating complex biological systems and for the further development of molecular tools, since imaging will be necessary to elucidate processes in their full complexity across scales. The technical resources within the CBCF also provide an ideal platform for the development and optimisation of small-molecule and protein-based molecular tools to study biological systems using imaging. The systematic investigation of their structure–property relationships in tailored assays, using high-throughput mutagenesis and chemical synthesis of small-molecule imaging reporters, will result in improved labels for fluorescence microscopy and novel imaging modalities.

Finally, it's possible to generate high-quality validated **small-molecule probes** that perturb the function of individual proteins involved in biomedically relevant processes. This can help to establish the potential of these proteins as therapeutic targets. In human health, systematic approaches are needed to identify validated chemical probes that can be matched to specific human or bacterial proteins. This makes it possible to elucidate links between target and phenotype for diseases of great unmet medical need (Figure SS6). The Open Targets platform at EMBL-EBI, which integrates many disparate data types for drug target identification and prioritisation, could serve as a starting point to identify proteins that could benefit the most from the development of specific molecular probes.

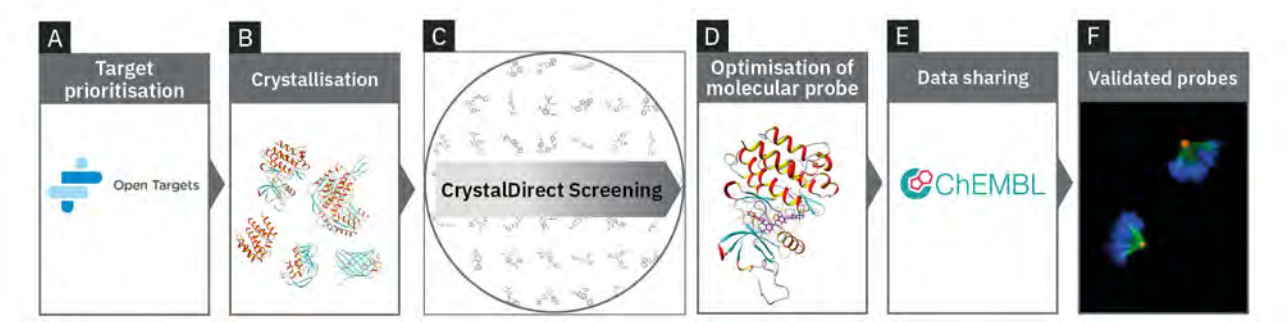


Figure SS6 | Systematic workflow to identify, design, develop, and validate high-quality molecular probes.
(A) Prioritisation of proteins for crystallisation using Open Targets. **(B)** Target proteins are expressed and robust crystallisation conditions are identified. **(C)** The targets are then screened against compound fragment libraries using the CrystalDirect (high-throughput crystallography) platform to identify binding compounds. **(D)** Structure-aided chemical optimisation, in combination with biophysical techniques, e.g. surface plasmon resonance, are used to develop a validated tool compound shown co-crystallised with its target protein. **(E)** Optimisation datasets are made publicly available via ChEMBL. **(F)** These specific validated tool compounds, matched with negative controls, can then be used to probe biological function.

Biological activity and structural data generated by the CBCF in collaboration with core facilities across EMBL will feed back into the public data resources at EMBL-EBI and provide a rich resource for researchers in EMBL's member states and beyond. Addressing global challenges such as climate change, environmental pollution, and threats to human health requires systematic innovation and the translation of research results from academia to industry for the benefit of society. Therapeutically relevant discoveries can also be exploited to develop novel drug candidates from the initial probes for the benefit of human, animal, or planetary health.

Mobile Services

One focus of the EMBL Programme will be on Planetary Biology, with a central goal to understand ecosystems at the molecular level. A key tenet to the Planetary Biology theme will be the development and extensive use of mobile lab services (Figure SS7). These mobile services will offer technologies for sampling and measurement, and will support longitudinal observational and other projects by scientific partners across EMBL member states. One example of such a project is TREC (Chapter 7: Planetary Biology). In cooperation with partners, this flagship project aims to sample and study organisms and the environment along land-water interfaces, with the goal of obtaining a molecular understanding of the relationships between various environmental factors – including pollutants – on microbial communities, as well as on key organisms about which EMBL has expertise.

Cutting-edge technology, as it becomes available or is developed by EMBL, will be required for the mobile lab services, together with the ability to record environmental parameters. This technology includes advanced imaging, complex molecular profiling, and advanced data management, integration, and analysis. Technologies will be set up in dedicated sampling vehicles and standardised containers to serve and support teams sampling ocean, coastal, and terrestrial regions and allow field experimentation. The containers can be customised so that collaborators can host technology and infrastructures complementary to those available at a particular site.

This project will help to establish shared, standardised local experimental platforms, coupled with global data storage, handling, analysis, and modelling workflows. As well as research and services, the mobile labs will also train and engage with scientists, students and the public, providing multiple benefits for the member states (Chapter 11: Training; Chapter 15: Public Engagement, Communications, and Outreach). For example, there will be opportunities to train local scientists in the advanced technologies offered by the mobile services. Local scientists will support the development of skills and infrastructures at partner institutions, as required for local longitudinal sampling projects.

Figure SS7 | Rendering of the interior of a possible mobile lab.

Molecular Data Services

EMBL is a world leader in big data bioinformatics, hosting and running premier biological databases at EMBL-EBI. Research datasets submitted to EMBL-EBI deposition databases (also known as archives) are processed, curated, and integrated to be findable, accessible, interoperable, and reusable (FAIR principles). The public data resources hosted by EMBL-EBI support reuse of many forms of data, including DNA sequences (genes and genomes), RNA sequences, information about proteins and macromolecular structures, cell metabolites, biological images, phenotypic data, and data on the effects of drugs on cells and tissues. Over 40 biomolecular databases are currently hosted by EMBL-EBI, its data resources were accessed more than 80 million times a day by the end of 2019, and EMBL-EBI's raw data storage capacity is more than 300 petabytes.

The open and global sharing of this information with the life sciences community has been foundational for multiple data-driven discoveries, providing a practical understanding of health and disease and an increased comprehension of biodiversity and the living environment. The deposition of specific types of research data

into single repositories has enabled powerful meta-analyses to be carried out on the combined datasets. Bigger and broader datasets enable more far-reaching questions to be asked, and make it possible to observe both large and small effects.

The technological advances in data generation also demand innovative developments in bioinformatic data services and research methods. EMBL-EBI implements robust and responsive software development practices with website designs and search systems that enhance usability of data. To support the growth of bioinformatics and meet the demands of the scientific community, the new EMBL Programme will drive scalable and robust data strategies that will enable accelerated science and further the use of open data in other scientific fields. EMBL-EBI will continue to provide the core worldwide storage for all genomics data, as well as deepening and further integrating the fundamental information infrastructures around human and medically relevant model organisms, from genomes to protein structures. In line with the new scientific directions in this Programme, and in collaboration with numerous partners worldwide, EMBL will also drive greater usage, application, and coordination of bioinformatics resources and tools in the following priority areas.

Bioimaging Reference Data, Standards, and Tools

Advances in imaging at all levels, made possible by revolutionary technologies including cryo- and volume electron microscopy, super-resolution and light-sheet microscopy, and X-ray imaging are enabling the generation of vast quantities of high-resolution imaging data. To deliver on the promise held out by imaging technology, it is essential that scientists openly share biological image data. This need led to the launch in 2019 of the EMBL-EBI **BioImage Archive** for reference bioimage data (Figure SS8). The BioImage Archive and the Imaging Centre in Heidelberg are part of a wider EMBL drive to improve access to imaging technology. The BioImage Archive, which continues to be developed, serves as the central bioimage data repository for the scientific community, while added-value databases consume reference datasets and enable reuse and integration. Reference images that are freely available for reuse include any image that has been formally published, as well as other, curated image datasets.

The archive will support a host of value-added image data resources, which will enhance the scientific value of the archival images through curation, integrative analysis, and the development of new analytical methods. Efforts are well under way to connect to EMBL-EBI data resources such as the **Electron Microscopy Public Image Archive (EMPIAR)**, for processed EM data, and **BioStudies**, the data resource for data that do not fit in the structured archives at EMBL-EBI. In the future, the BioImage Archive will serve as the foundational archive supporting multiple user community-developed databases, e.g. for plant phenotypes, enhancing the scientific value of the archived images through curation and the development of new analytical methods. The archive and corresponding coordinated data resources are critical for new opportunities in research, development of new methods, and training. As well as connecting to the sites across EMBL generating imaging data (Chapter 8: Data Sciences), EMBL-EBI will continue to engage with other high-throughput biomolecular data-generation institutes in strategic collaborations to develop data-harvesting and deposition pipelines and metadata standards that work across the various imaging communities.

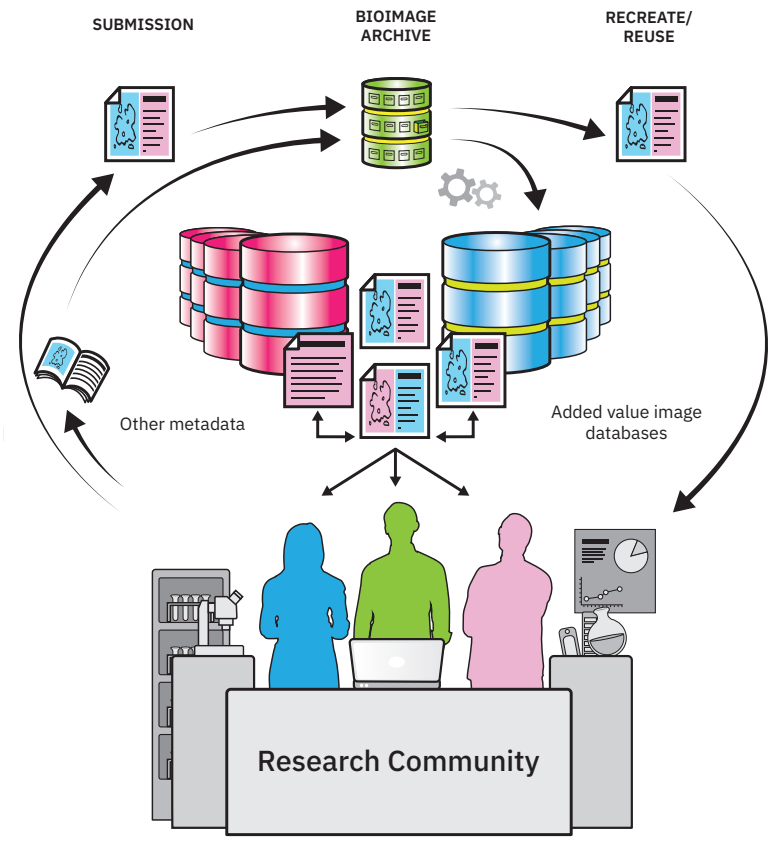


Figure SS8 | The BioImage Archive.
The flow of data between different parts of the community is shown. The BioImage Archive serves as the central bioimage data repository for the scientific community, while added-value databases consume reference datasets and enable reuse and integration.

Genomic Medicine Platform

Genomic medicine is the use of genomic information or other genome-wide molecular measurements in the context of clinical practice. This includes the sequencing of germline DNA for individuals with suspected rare genetic diseases and the sequencing of somatic cancer DNA and RNA from individuals with cancer, as well as other diseases. A number of countries have implemented one or more of these strands, with a substantiated positive impact on healthcare for their citizens. An EMBL study estimates that over 60 million patients will have their genome sequenced in a healthcare context by 2025. In some cases, there is also evidence for cost savings within a country's healthcare system.

In the next decade, genomic medicine is likely to enable risk stratification for a variety of diseases by integrating whole-genome information, sequencing the human microbiome in a variety of niches, sequencing RNA (potentially at the single-cell level), and profiling the immune repertoire in immune-related diseases, including cancer. Many aspects of medical practice can be enhanced, sometimes transformed, by the integration of genomic medicine. Over the past five years, EMBL has played a role in supporting and enabling a number of EMBL member states to make substantial progress in designing, delivering, and operating genomic medicine services. This includes, for example, the provision of both strategic and technical support to the UK 100,000 Genomes Project and the Danish Genomic Medicine service. EMBL is also a founding member of the Global Alliance for Genomics and Health (GA4GH) (Chapter 6: Human Ecosystems), the international coalition formed to enable standards for the responsible sharing of genomics data in healthcare.

In the next EMBL Programme, EMBL will implement a bespoke, multi-component service, the **Genomic Medicine Platform**, which will engage with individual national initiatives, with EMBL scientists advising and proactively transferring technology into EMBL member states that have embarked on bringing precision medicine into their healthcare systems (Figure SS9). This scheme will take advantage of multiple developments by EMBL's bioinformatics teams in developing data structures and standards that promote: (1) the proactive sharing of technology from EMBL with member states, (2) the provision of key reference knowledge bases for interpreting genomes, and (3) the responsible secondary use of genomics data for research, with systems that respect national legal and cultural requirements, and associated processes. The Genomic Medicine Platform has five main arms:

1. **Strategic advice**: With experience and evidence from various international settings, EMBL can provide scientific and technical advice on the organisation of effective bioinformatics systems and services needed for the large-scale deployment of national initiatives, always mindful of the legislation and operational framework of each EMBL member state.
2. **Reference human molecular data**: Via its open data repositories like **Ensembl**, **UniProt**, and **Open Targets Platform**, EMBL-EBI provides the reference human genome sequence, reference gene annotation such as gene locations and protein function, and links to protein structure, pathways, expression, metabolism, and disease information. Tools such as the **Variant Effect Predictor** can be used to interrogate clinically observed variation against this reference information.
3. **EGA Federation**: In collaboration with the Centre for Genomic Regulation (CRG) in Spain, EMBL has developed 'federated access' models of the European Genome-phenome Archive (EGA) data resource, which holds data from human studies with research usage conditions based on consent permissions in the datasets. EMBL will provide strategic and operational support in establishing federated EGA deployments that involve back-end design, software, policy development, and installation, as demonstrated in the pilot study to establish the German Human Genome-Phenome Archive.
4. **Technical training courses**: To enable further knowledge transfer in this area, EMBL will run bespoke technical training courses on informatics methods and standards in EMBL member states by invitation. This is best done once the delivery plan for genomic medicine in the EMBL member state is set.
5. **Mid- and long-term secondments**: EMBL will offer mid- to long-term secondments to enable staff from EMBL member state organisations to work alongside EMBL-EBI staff. Secondments will be focused on genomic medicine engineering and data management techniques to maximise technology and skills transfer into each member state.

As this science progresses, EMBL aims to be a neutral and fair-sharing portal of knowledge across Europe and the world, allowing every country to maximise the delivery of accurate genomic medicine, and in particular ensuring that EMBL member states have the latest and most accurate information possible. EMBL's involvement in the EU 1+ Million Genomes initiative is a potential route by which genomic medicine services could be provided to the 22 EU states that are part of the initiative.

Figure SS9 | Genomic Medicine Platform.

EMBL's new multi-component service will proactively transfer knowledge and technology into national precision medicine initiatives in EMBL member states through the five main components illustrated.

Human Brain and Behaviour Data

A major new service activity will be developed as part of the Centre for Human Brain Phenomics at EMBL Rome (Chapter 6: Human Ecosystems). A major aim of the centre will be to develop data-mining and analysis tools for human big data related to brain function, and make these tools available to the research community. Big data from human brain imaging (rs-fMRI, EEG, MEG, PET) and studies of behaviour (social media, wearable biosensors, CCTV data) are complex, and analysis tools and standards are only now emerging to perform the required dimensionality reduction, visualisation, and statistical assessment needed to extract genotype–phenotype links. However, most researchers are not equipped to handle or analyse such datasets. This lack of expertise, combined with a lack of infrastructure, means that investments in research at the national level are often ineffectual and disparities in research infrastructures across European countries are accentuated. The Centre for Human Brain Phenomics will satisfy this need by leveraging EMBL's unique expertise in data sciences (Chapter 8: Data Sciences) with cutting-edge research in statistical modelling and image analysis to develop novel tools that will allow member state researchers to successfully handle and analyse brain-related big data. The centre will also be able to enter into collaborative partnerships with national research consortia to facilitate big data access and to provide data services in the area of human brain and behaviour big data research.

Genes, Genomes, and Variation

EMBL's flagship project TREC (Chapter 7: Planetary Biology), and participation in initiatives such as the international Earth BioGenome Project (an effort to sequence the genomes of all species on Earth) and the associated Darwin Tree of Life project (to sequence the estimated 80,000 eukaryotic species in the British Isles), will generate a vast amount of genomic and protein data and will aid the understanding of evolutionary mechanisms and diversity. Meanwhile, environmental sampling (which negates the need for the isolation and culturing of organisms) has demonstrated that the majority of species (approximately 99%) have not yet had their genomes sequenced, highlighting the huge untapped genetic potential waiting to be discovered.

The **Ensembl** data resource is used for the creation, integration, and distribution of reference datasets and analysis tools for genomics research. Data resources like Ensembl have to provide rich, deep, and scalable systems to handle data, allowing the genomes of every species to be stored, analysed, and compared, with the resulting information easily visualised or made accessible through APIs or downloads. Organising the data from thousands of genomes will make it possible to capture a measure of global biodiversity at an unprecedented level of detail, allowing scientists to trace lineages of evolutionary change, manifested at either the level of the DNA sequence or the genes encoded, which may be correlated with functional traits or adaptations to a particular habitat. This depth of genomic sequencing will also be cross-cutting to other data services. For example, the data will enable the identification of deleterious mutations and will facilitate the structural modelling of proteins through the application of co-evolutionary approaches. Such models, captured in **PDBe**, make it possible to predict the function of previously uncharacterised proteins, and provide clues as to how high-resolution structures may be obtained via other EMBL services.

The diminishing costs of generating DNA sequences are also transforming the way we understand communities. For example, metagenomics – the analysis of the sum of DNA found within a sample – is increasingly being used to understand the microbial component (the microbiome) of a diverse range of samples from the human gut to the deep oceans to agricultural soils. To deal with this burgeoning range of datasets, enhanced archiving and data analysis capabilities will be developed for **MGnify**, the EMBL-EBI repository and analysis service for metagenomics and metatranscriptomics of microbiome data. Of particular note will be the assembly and reconstruction of near-complete genomes (viruses, bacteria, and single-cell eukaryotes), which represent the vast majority of genomic space yet to be sequenced and exploited. However, as so much of this data is novel, new data resources that provide functional annotations and exploratory tools of the microbiome proteome will also be developed (Chapter 4: Microbial Ecosystems). Bioinformatics tools and data resources that can help identify similarities and differences between these organisms, as well as the details of their symbiotic relationships, will be crucial in understanding the interplay and interaction we see in our environment.

In addition to understanding microbes, similar molecular techniques are increasingly being used at a global scale to examine the spatial and temporal distribution of larger organisms (e.g. algae, fungi, insects, plants, and vertebrates) via traces of their environmental DNA (eDNA). EMBL will extend the MGnify pipelines to enable the capture and analysis of eDNA, to provide another resource for assessing biodiversity and to link the occurrences of species to the genomic records in Ensembl. This will enable the modelling of complex ecosystems at the molecular level. The standardisation and archiving of species observation records, as offered by MGnify, will enable scientists to monitor everything from individual species to overall biodiversity. Such distributed datasets are sourced from sampling stations such as the Global Omics Observatory Network (GLOMICON) and its partners in the Genomic Observatories Network. This species data then flows to other resources, such as the Global Biodiversity Information Facility (GBIF), and is used to guide governmental policies. These ambitious global projects provide further opportunities for EMBL to continue its role as a trusted neutral intermediary, and to promote data sharing and coordination between unconnected communities.

Multi-omics Data Resources

The predicted increase in quantitative transcriptomics, proteomics, phosphoproteomics, and metabolomics data will provide both richer datasets and more consistent surveys of transcriptomes, proteomes, and metabolites. EMBL-EBI's involvement in global data-generation projects often involves challenging aspects in terms of scale, complexity, and changes in technology. For example, the Human Cell Atlas project demanded an entirely new model for storing, processing, and analysing single-cell data on a large scale. The development of single-cell RNA-seq technologies and increased data submissions have driven the need for new infrastructure models and new methods to integrate RNA-seq with imaging (spatial transcriptomics). The challenges include handling data from millions of cells in a single experiment, dealing with novel definitions

of cells and tissues, and the development of a data coordination platform as a specialist secondary portal. The integration of imaging and genomics data in spatial transcriptomics, proteomics, and metabolomics will lead to increased complexity in the organisation and visualisation of cells and tissues. These technology developments will impact archiving and analysis services provided by EMBL-EBI data resources such as the **Single Cell Expression Atlas** (catalogue of single-cell expression data from a range of species), **PRIDE** (**PRoteomics IDEntifications Database** – the public data repository for proteomics data, including protein and peptide identifications, post-translational modifications, and supporting spectral evidence), and **MetaboLights** (database for cross-species and cross-technique metabolomics experiments and derived information). EMBL also aims to increase the provision of biologically relevant chemistry resources such as **ChEMBL**, which contains data from endogenous molecule identification through to synthesised chemical biology, including drug-like molecules.

Access to Research Publications and Associated Data

The means by which research articles are published continues to evolve as practices such as preprint publishing and sharing peer review materials increase in the life sciences. As new publishing platforms, open access policies, and data-sharing policies continue to develop, there is a growing need for public infrastructure that supports these changes. Open access to research publications is important not only because it enables users to freely access individual research articles, but also because of the programmatic access it provides to a collection of millions of papers. This makes it possible to gain further insights by analysing this collection using machine learning and by developing text-mining applications, helping to build the search and retrieval tools of tomorrow.

In the context of open science, cross-linking open access publications to the underlying open data is of fundamental importance. Equally, addressing the matter of unambiguous attribution of authorship, funding, and institutional affiliation will give rise to better mechanisms for assigning credit, which is a key part of science culture. Resources such as **Europe PMC**, the European hub for open research publications, and **BioStudies**, the resource that links together all the data behind a paper and stores unstructured (supplemental) data, actively support open science and literature–data integration, aligning with the open science policies of Europe's funding agencies.

Data Coordination via Data Portals

With bioinformatics now a mainstream discipline and crossing over into other disciplines such as engineering or medicine, the number of scientists who use biomolecular data in their research and the diversity of their academic backgrounds have increased substantially. The roles played by scientists can range from data producers and data submitters through to data consumers, those who analyse and interpret the data, as well as teachers, students, and trainers of bioinformatics (Chapter 8: Data Sciences). With the increasingly large quantity and intensive use of data in the life sciences, effective delivery of both scientific data and training to the worldwide scientific community are vital. New data types, data from multiple domains, and demands from diverse user communities to further their understanding of data, pipelines, and tools that connect various resources all mean that there is a need to build biology-specific portals. EMBL's delivery of such infrastructure for data sharing and collaborative analysis played a vital role in the rapid coordination of data during the COVID-19 pandemic, via the provision of the European COVID-19 Data Platform (Chapter 5: Infection Biology).

In this Programme, EMBL aims to increase its involvement in landmark global projects and consortium-led collaborations. These projects will drive major developments in bioinformatics, bringing together communities

to agree on data standards, coordinate data acquisition and sharing, support complex data flows, build repositories for data from emerging technologies, and develop infrastructure to serve scientific communities with diverse and often unpredictable needs. With collaborators, EMBL will support the development of biology-specific data portals. These will be single integrated points of entry, allowing scientists to understand the scale and complexity of the relevant data and facilitating the compelling and practical presentation of data. This will ensure easy access for both the general scientific user base and for scientists involved in these initiatives, and will allow external experts to build freely upon EMBL initiatives.

Conclusion

The future strength of EMBL's scientific services will come from the meaningful and increasingly seamless integration of complementary experimental and data services to understand the complexities of biological structure and function. EMBL, with its distributed and complementary activities in molecular biology, is in an excellent position to face these challenges. By synergising and leveraging its strengths in both experimental and data technologies and its longstanding experience in operating research infrastructures, EMBL can develop bilateral and multilateral interactions to motivate strong collaborative research and service development activities. This unique ensemble of expertise and approaches will push forward the life sciences in Europe.