# Appendix I: Guide to Technologies

This guide provides brief descriptions of the technologies referred to in the EMBL Programme. Some of these were developed by EMBL scientists and engineers. The technologies below cover structural biology and imaging, computational methods, omics technologies, and combinations of these.

# Guide to Technologies at EMBL

## Computational Technologies

**Cloud technology** enables the practice of cloud computing, which uses a network of remote servers hosted on the internet or in local data centres to store, manage, and process data.

**Deposition databases** (data archives) are used by researchers to submit their experimental data and supporting metadata, often at the time of publication. Deposition databases thereby preserve the scientific record and enable reuse of research data by others in the scientific community.

**Knowledge bases** (added-value databases) contain carefully selected combinations of data, which are combined and enhanced by computational and curational processes. Unlike deposition databases, knowledge bases provide a non-redundant, consistent view of the available data.

> Biology-driven portals connect data from multiple domains and can help users within specific communities to make the best possible use of data and tools in a user-friendly manner.

**Machine learning** is the study of algorithms that perform a specific task without explicit instructions. These algorithms automatically improve with experience, by relying on patterns and inferences in the sample data they are given.

## Omics Technologies

### Genomics Technologies

**Genome analysis** is the identification, quantification, or comparison of genomic features. This includes analysing genome sequences, structural variation, gene expression, or regulatory or functional element activity at a genomic scale.

> Next-generation sequencing (NGS) is a catch-all term used to describe a number of high-throughput technologies for determining the order of nucleotides in DNA or RNA that employ massive parallel sequencing of spatially separated, clonally amplified DNA or RNA templates or single DNA or RNA molecules in a flow cell.
>
> > RNA sequencing is the use of NGS to sequence and quantify the total complement of RNAs (the transcriptome) in a given sample, including methods to detect transcription in single cells (scRNA-seq).
> >
> > DNase I hypersensitive sites sequencing (DNase-seq) is a method to identify the location of regulatory regions of DNA, based on genome-wide sequencing of regions sensitive to

cleavage by the endonuclease DNase I. In these regions of the genome, chromatin has lost its condensed structure, exposing the DNA and making it accessible for the binding of proteins such as transcription factors.

Assay for transposase-accessible chromatin using sequencing (ATAC-seq) is an alternative method to map the location of regulatory regions genome wide, by probing open chromatin with hyperactive transposase Tn5, which inserts DNA tags (sequencing adapters) into open regions of the genome.

Chromatin immunoprecipitation sequencing (ChIP-seq) combines chromatin immunoprecipitation (ChIP) with NGS to identify and map the genomic DNA regions that bind to DNA-associated proteins (such as transcription factors, chromatin-modifying enzymes, modified histones, and components of the basal transcriptional machinery).

**Chromosome conformation capture (3C)** techniques are a set of molecular methods to analyse the three-dimensional organisation of genomic regions in the nucleus by quantifying the frequency of interactions between genomic loci, even those located very far from each other in the linear genome.

**Single-cell tri-channel processing (scTRIP)** is a technique to selectively sequence a daughter cell's parental template strands to reconstruct complete karyotypes of single cells, enabling systematic analyses of large structural variants including duplications, deletions, balanced inversions, and translocations.

**Combinatorial indexing** allows the labelling of chromatin in thousands of individual cells using Tn5 transposase loaded with different barcoding DNA templates. Cells are then pooled and sequenced, after which all reads can be deconvoluted to detect biological signals from each individual cell.

**Genome editing techniques** enable specific changes to be made to DNA by molecularly cutting the DNA at a specific sequence. When this cut is repaired by the cell, a change or 'edit' is made to the sequence to either modify or remove a gene or modulate its regulation.

Clustered regularly interspaced short palindromic repeats (CRISPR) genome editing allows genes (or regulatory DNA) to be removed or added into a genome using a Cas9 nuclease complexed with synthetic guide RNA to precisely locate and cut an organism's genome.

Adeno-associated virus (AAV) vectors are used for high-efficiency, cell-type-specific CRISPR-based gene knockouts in mice and tissue organoid systems.

**Degron approaches** enable rapidly inducible degradation of a target protein's native expression levels and patterns, allowing specific and temporally defined perturbation for any protein of interest.

## Proteomics Technologies

**Mass spectrometry (MS)** is an analytical technique that measures the mass-to-charge ratio of ions of one or more molecules present in a sample to identify unknown compounds via molecular weight determination, or to quantify known compounds, or to determine structures and chemical properties of molecules.

**Proteomics** is the large-scale study of proteomes, the set of proteins produced in an organism, system, or biological context. Several high-throughput technologies have been developed to investigate proteomes in depth, such as MS-based techniques (e.g. tandem MS) and gel-based techniques such as differential in-gel electrophoresis (DIGE).

**Thermal proteome profiling (TPP)** measures the heat-induced denaturation of proteins using quantitative MS. This can be used to identify drug binding properties of protein targets based on altered thermal stability.

> Solubility proteome profiling (SPP) broadly characterises the solubilising effects of a molecule (e.g. ATP) on the proteome.

**Multiplexed proteome dynamics profiling (mPDP)** is an MS-based approach combining dynamic SILAC (stable isotope labelling by amino acids in cell culture) with isobaric mass tagging for multiplexed analysis of protein degradation and synthesis.

### Metabolomics Technologies

**Metabolomics** is the large-scale study of metabolites (small molecules, lipids, drugs, microbiome products, and other exogenous molecules such as pollutants, pesticides, and toxins) and their transformations within cells, tissues, and organisms.

**Matrix-assisted laser desorption/ionisation (MALDI)** is an ionisation technique that creates ions from a broad range of molecules with high sensitivity and minimal unwanted fragmentation. Combined with mass spectrometry, MALDI can be used to detect and quantify metabolites.

> Single-cell metabolomics enables metabolites to be detected in individual cells, including rare cell types or states, such as stem cells or infected cells. When integrated with advanced microscopy, this technology can reveal metabolic reprogramming at the single-cell level.

> Spatial metabolomics by means of MALDI imaging mass spectrometry can be used to detect a broad range of metabolites in tissue sections at cellular resolution. The resulting metabolite images can be overlaid with microscopy images to link metabolite information with tissue morphology (H&E), expression of proteins (immunofluorescence, immunohistochemistry) and genes (*in situ* hybridisation).

## Chemical Biology

**Fluorophores** are fluorescent molecules that can re-emit light after absorption of light of a different wavelength. Fluorophores can be either synthetic (based on small molecules that are chemically synthesised), or genetically encoded (based on naturally occurring fluorescent proteins). They are the essential reporters that provide contrast in light microscopy. Their properties, such as colour, can be altered by introducing structural modifications into the molecule.

> Fluorogenic probes are molecules that are non-fluorescent but become fluorescent upon binding to a specific target. This ‘turn-on’ effect guarantees a high signal-to-noise ratio in fluorescence microscopy, as unbound molecules do not display any signal.

> Fluorescent indicators are molecules whose fluorescence signal is modulated in response to a specific analyte. They are used in fluorescence microscopy to report on dynamic processes such as fluctuations in pH, ion concentration, or the presence of a small molecule of interest.

> Photoresponsive fluorophores are molecules that fluorescence emission changes in response to light. Upon illumination, they undergo a transformation from an ‘off’, non-fluorescent state, to an ‘on’, fluorescent state.

**Labelling methodologies** encompass all approaches used to attach a reporter to a molecule of interest. These methods are necessary to be able to visualise specific features in biological systems with high contrast. The most commonly used labelling methodologies rely on the genetic fusion of a protein to a molecule of interest. This protein can either be intrinsically fluorescent, or can recognise and bind a synthetic fluorophore which is exogenously added.

## Structural Biology and Imaging Technologies

### Energy source: Photons

**Intravital imaging** is a thick tissue imaging technique to observe live biological processes at high spatial resolution, enabling non-invasive and longitudinal studies of living organisms. Fluorescence reporters are employed as a contrast agent and any imaging technique that provides intrinsic optical sectioning capability, such as confocal or multiphoton microscopy can be used.

**Single-photon microscopy (confocal microscopy)** is a classic fluorescence microscopy technique that focuses a point of light onto one plane within a biological sample. Fluorophores are excited and a longer wavelength of light is emitted, which can be detected. Out-of-focus emitted light is blocked by a pinhole, increasing optical resolution and contrast.

**Multiphoton microscopy** uses near-infrared (NIR) femtosecond pulses of laser light to generate nonlinear signals in the visible range. Unlike traditional single-photon fluorescence microscopy, the wavelengths of the two exciting photons are longer than the wavelength of the resulting emitted light, thus enabling deep imaging into thick tissue specimens.

**Light-sheet fluorescence microscopy (LSFM)** enables the imaging of delicate samples, such as dividing cells in an embryo. A sample slice is illuminated perpendicular to the direction of observation using a plane of laser light. Images can be acquired at speeds 100 to 1000 times faster than single- or multiphoton microscopy, while reducing the photodamage and stress inflicted upon a sample.

**Super-resolution microscopy** covers a range of microscopy techniques that enable images to be taken at a resolution higher than the diffraction limit.

> Single-molecule localisation microscopy (SMLM) is dependent on the switching of a single fluorophore from an ‘off’, dark state, to an ‘on’, emission state, before photobleaching to ‘off’. This process is sequentially repeated until most fluorophores have been imaged. Sparse activation allows for precise determination of the positions of single fluorophores. Many images are then reconstructed into one, enabling individual molecules to be precisely located.
>
> Stimulated emission depletion (STED) creates super-resolution images by the selective deactivation of fluorophores in specific regions of a sample while leaving a central focal spot active to emit fluorescence. Minimising the area of illumination at the focal point enhances the achievable resolution for a given system.
>
> MINFLUX combines single-molecule activation (as in SMLM) with targeted scanned read-out using a doughnut-shaped beam (as in STED) to localise single fluorophores with nanometre resolution. Targeted activation can greatly increase the temporal resolution for live-cell imaging.

3D super-resolution microscopy encompasses several techniques that enable the three-dimensional visualisation of biological samples with super-resolution.

Astigmatic SMLM uses a cylindrical lens in the detection path to introduce astigmatism, i.e. a strong *z*-dependent deformation of the single-molecule images, which are analysed to obtain precise *z*-positions.

Supercritical-angle localisation microscopy evaluates the near-field emission of single fluorophores together with their far-field emission in a dual-channel setup to obtain absolute *z* coordinates with unprecedented accuracy in the vicinity of the cover slip.

4Pi-SMLM uses four-channel interference of single fluorophore emission to extract highly precise *z* positions from the interference phase, even in thick samples.

3D STED employs two phase masks to create a hollow three-dimensional de-excitation volume to confine the emission not only in the lateral, but also in the axial dimension.

3D MINFLUX is an extension of MINFLUX using a hollow detection volume that enables single nanometre isotropic 3D resolution, even in live cells.

**Mesoscopic imaging** is the high-resolution imaging of biological systems from millimetre- to centimetre-sized biological samples, such as tissues, organs, or small organisms.

Optical projection tomography (OPT) illuminates and collects wide-field (fluorescence) images from multiple angles to allow tomographic image reconstruction in post-processing. OPT can enable rapid mapping of the tissue distribution of RNA and protein expression in intact embryos or organ systems.

Selective plane illumination microscopy (SPIM) generates high-resolution, optically sectioned, multidimensional images by combining 2D illumination with orthogonal camera-based detection. This is achieved with minimal photodamage and at speeds capable of capturing transient biological phenomena.

**Photoacoustic tomography** uses short, high-energy laser pulses tuned to the absorption spectrum of molecules of interest for ultrasound-based imaging. Since sound waves are significantly less prone to scattering than light, this technology enables imaging of centimetre-sized samples at mesoscopic resolution.

## Energy Source: Electrons

**Cryo-electron microscopy (cryo-EM)** uses a transmission electron microscope at cryogenic (liquid nitrogen) temperature to visualise vitrified specimens (frozen but without the formation of crystalline ice). Structures of interest within biological samples can be visualised in the context of their native cellular or tissue environment to yield 3D structures of protein assemblies (macromolecular complexes or organelles) at near-atomic resolution.

Cryo-electron tomography (cryo-ET) is a process in which a series of images are taken of a sample tilted at different angles with respect to the electron beam, producing a series of 2D images that are then combined into a 3D reconstruction (tomogram). Computing methods (subtomogram averaging) are then used to generate high-resolution maps.

Single particle cryo-EM is the analysis of several combined, digitised cryo-EM images of similar molecules imaged in multiple orientations. This analysis produces image averages, which reduce noise and help to visualise interpretable structural features, allowing a 3D atomic model of a molecule to be built up.

**Volume scanning electron microscopy (volume SEM)** is a family of techniques that automatically generate stacks of images from serialy cutting bulk specimens. The volumes generated are typically used to assess connectivity and 3D organisation of organelles and cells at subcellular resolution.

Serial block-face electron microscopy (SBEM) generates high-resolution 3D images from small resin-embedded samples. Thin sections from the face of the block are cut and the sample is imaged again, enabling many thousands of aligned images to be acquired in an automated fashion.

Cryo-focused ion beam scanning electron microscopy (cryo-FIB-SEM) is generally used to create electron-transparent lamellae through cells or micromachine them out of tissues. Such lamellae are then transferred to a transmission electron microscope for cryo-ET.

## Energy Source: Radio Waves

**Nuclear magnetic resonance (NMR) spectroscopy** is a structural biology method that involves the observation of local magnetic fields around atomic nuclei. Liquid or solid samples are placed in a strong magnetic field in which the NMR signal is produced by excitation of the nuclei with radio-frequency pulses. The signal reports on the chemical environment of all atoms in the sample. This enables NMR spectroscopy to provide detailed information about the identity, structure, dynamics, reaction state, and interactions of biological macromolecules at atomic resolution.

## Energy Source: X-rays

**X-rays** can be produced by electrons accelerated to almost the speed of light in large-scale machines such as synchrotrons and X-ray free-electron lasers, which are then guided to experimental stations in beamlines. Due to their high intensity, short wavelength, and large penetration depth, X-rays play an important role in high-resolution structural biology and imaging technologies.

**Small-angle X-ray scattering (SAXS)** is used to determine the structure of proteins or macromolecules without the need for crystallisation. This is achieved by recording and analysing the scattering behaviour of X-rays at small angles when travelling through a solution of molecules.

**X-ray crystallography of proteins (PX) or macromolecules (MX)** is used to determine the crystal structure of these molecules by measuring the angles and intensities of X-rays diffracted by their crystals. 3D structural models describing the positions of individual atoms can be determined from diffraction data.

Time-resolved pump-probe serial crystallography enables the physical movements of a protein within a crystal to be visualised. Crystals containing protein molecules and ready-to-go substrate molecules are pumped – often with a laser – to initiate the chemical reaction. The structure is then probed with an X-ray pulse at varying times after the chemical reaction was initiated, to measure the protein conformation at each time point. To obtain full 3D information for each time point, large series of crystals must be interrogated. Datasets can then be assembled to show the movement of the protein structure.

High-energy X-ray phase contrast tomography (nanoholotomography) is a process in which a series of images is taken of a sample rotated with respect to the X-ray beam. The resulting series of 2D projection images is then combined into a 3D tomographic reconstruction, showing the shape and inner organisation of the sample at sub-micrometre resolution.

High-throughput X-ray tomography allows for large-field-of-view, large-scale screening of biological samples, enabled by the short timescale for completing measurements and the automation of data collection and analysis processes at the endstation of a beamline.

A **diffractometer** is a high-precision mechanical instrument for precisely rotating micrometre-sized crystals in a micrometre-sized X-ray beam to record the resulting diffraction patterns. A high-throughput diffractometer, **CrystalDirect**, was developed at EMBL and enables the screening of thousands of crystals at beamlines worldwide.

### Biophysical Technologies

**Fluorescence correlation spectroscopy (FCS)** is used to measure concentrations and diffusion coefficients of fluorescently labelled molecules. It can be combined with other imaging modalities. Light is focused on a sample in a small imaging volume, and the measured fluorescence intensity fluctuations due to diffusion or physical or chemical reactions are analysed using temporal autocorrelation.

**Atomic force microscopy (AFM)** uses nanometre- to micrometre-sized cantilevers to measure the mechanical properties of a sample by physically exerting a force onto the material while recording its viscoelastic response, to infer biophysical properties of membranes, cells, or entire tissues.

**Laser ablation** is the process of destroying biological material by using high-intensity, ultrashort laser pulses. This enables cutting of biological structures, such as actin filaments, in a spatially and temporally controlled manner. The method is used to probe the viscoelastic response of tissue components, induce apoptosis, or ablate nuclei to measure the dynamic tissue response upon ablation in spatio-temporal detail.

**Magnetic droplets** are micrometre-sized biocompatible ferrofluid oil droplets that can function as local mechanical actuators and allow quantitative, spatio-temporal measurements of mechanical properties in 3D and *in vivo*. These droplets can be injected or otherwise embedded between cells inside a tissue, with magnets exerting controlled forces that lead to droplet deformations. By imaging both the response of the droplet and its immediate surroundings, the mechanical properties of the tissue can be inferred.

**Brillouin microscopy** measures the mechanical properties of biological samples, such as elasticity and viscosity, in a non-contact and label-free fashion in 3D with high, diffraction-limited resolution. Laser light is focused on a sample, where it interacts with sound waves that are intrinsically present in any material because of thermal agitation. During the scattering process, the light experiences a positive or negative frequency shift, the amount of which is proportional to the speed of sound inside the material and therefore provides information on its viscoelastic properties.

## Multimodal Approaches

Multimodal approaches measure multiple modalities in one experiment or integrate datasets from multiple experiments.

**Integrative structural biology** combines highly complementary structural biology techniques such as NMR, X-ray crystallography, small-angle X-ray scattering, and cryo-EM.

**Correlative multimodal imaging (CMI)** integrates multiple imaging modalities such as light microscopy, (cryo-)electron microscopy, or X-ray imaging to link structural and functional analyses at multiple scales.

> Correlative light and electron microscopy (CLEM) integrates light and electron microscopy techniques for a single sample. Development of hardware, for example to work at cryogenic temperatures, and of software constantly push the boundaries of the technique for a wide range of applications.
>
> Optogenetics is a combination of optics and genetics to control the activities or location of individual molecules inside living cells. Light can be used to control single cells (including in living tissues or embryos) that have been genetically modified to express light-sensitive proteins.

**Chemogenetics** is a combination of chemical biology and genetics to control the activity of molecules inside cells. Small molecules and drugs can be used to control single cells (including in living tissues) that have been genetically modified to express chemosensitive proteins.

**Integrated single-cell multi-omics** combines various single-cell genomics technologies to measure different types of molecular data from the same cell. For example, measuring gene expression and epigenetic signatures, or the genotype (DNA) and expression state (RNA) of individual cells.

**Integrated single-cell spatial omics** combines various omics technologies with microscopy images to gain a spatial understanding, at single-cell resolution, of different types of molecular data within a biological sample. This can be used, for example, to measure the expression of all genes within each cell in a tissue, while preserving spatial information about the positions of cells relative to one another.

> Sequential fluorescence *in situ* hybridisation (seqFISH) enables the identification of thousands of RNA transcripts within single cells via sequential rounds of hybridisation, while maintaining spatial information about the tissue or embryo.