# **Project Title:** Development of Interactive Visualisation and Conversational AI Tools for Spatial Transcriptomics Data

**Primary Supervisor:**
Christina Ernst (EMBL-EBI)

**Co-supervisor(s):**
Julio Saez-Rodriguez (EMBL-EBI), Muzlifah Haniffa (Sanger Institute)

## Project Description:

### Background:

Technological advances in single-cell and spatial transcriptomics have significantly enhanced our understanding of cellular heterogeneity and spatial organisation within tissues. Projects such as the Human Cell Atlas (HCA) have generated extensive datasets with millions of cells mapped from diverse tissues. Despite the immense potential, the complexity of these datasets often restricts access to researchers with advanced computational skills and resources, thereby limiting broader scientific and clinical utility.

Current platforms for visualisation and exploration of spatial transcriptomics data remain fragmented and limited in functionality, scalability, and accessibility. Additionally, they do not yet leverage the powerful potential of interactive conversational AI for data querying and exploration. Developing tools that democratise access to these datasets, allowing intuitive interaction and visualisation, is crucial to unlocking their full potential.

### Project Outline and Aims:

This project aims to develop a novel, robust visualisation platform integrated with conversational AI tools to facilitate interactive exploration of spatial transcriptomics data. Specifically, the project will:

1. Develop advanced interactive visualisation functionalities for spatial transcriptomics data, leveraging existing tools (WebAtlas, Vitessce, Zarr, SpatialData).
2. Integrate innovative conversational AI models to allow intuitive querying and interactive dialogue with datasets, significantly reducing barriers to data exploration.
3. Ensure optimal data accessibility and performance through scalable data delivery and lightweight visualisation strategies.
4. Implement ontology-driven, multilingual search capabilities, enhancing global accessibility and data discoverability.

The Fellow will design and prototype web applications using WebAtlas and Vitessce frameworks to support interactive visualisation of spatial datasets (e.g., 10x Visium, Xenium). If necessary, the project will involve extending these frameworks or data models to accommodate novel spatial transcriptomics technologies, analytical methods, or visualisation strategies.

The project also involves incorporating conversational AI tools using large language models (LLMs) such as those based on the BioChatter framework, with integration of ontology-aware knowledge graphs. These will support natural language interfaces for querying datasets by gene, cell type, tissue, or condition, and provide semantically informed responses that can guide users through available data, suggest relevant subset for analysis, and explain observed patterns based on embedded metadata and derived features.

In addition to bridging research groups at EMBL-EBI and the Sanger Institute, this project also involves close integration with EMBL-EBI's Functional Genomics Team, which provides long-standing services for the curation and dissemination of transcriptomic data. This unique alignment ensures that the tools and methodologies developed by the Fellow will not only contribute to academic research, but also directly enhance EMBL-EBI's service portfolio. By embedding the outcomes into existing submission, processing, and visualisation infrastructure, the project has the potential to significantly broaden access to spatial transcriptomics data, ensuring long-term usability and sustainability beyond the fellowship period.

**Training and Mentorship:**

The Fellow will gain interdisciplinary experience spanning bioinformatics, spatial biology, data visualisation, and AI-driven tool development. They will receive hands-on training in:

- Spatial transcriptomics data processing and annotation pipelines
- Scalable web application development for high-dimensional data
- Implementation of natural language interfaces using LLMs and ontology-based search
- Collaborative tool development across wet and dry lab environments

This project is co-supervised by Christina Ernst and Julio Saez-Rodriguez at EMBL-EBI, and Muzlifah Haniffa at the Sanger Institute, offering a unique opportunity to work across complementary research areas at the intersection of computational biology, AI, and human tissue atlases. This collaboration also exemplifies the potential for high-impact, open science development across research and service teams, strengthening the foundation for sustainable, community-accessible tools in spatial biology.