ICR | The Institute of Cancer Research

NIHR | Biomedical Research Centre at The Royal Marsden and the ICR
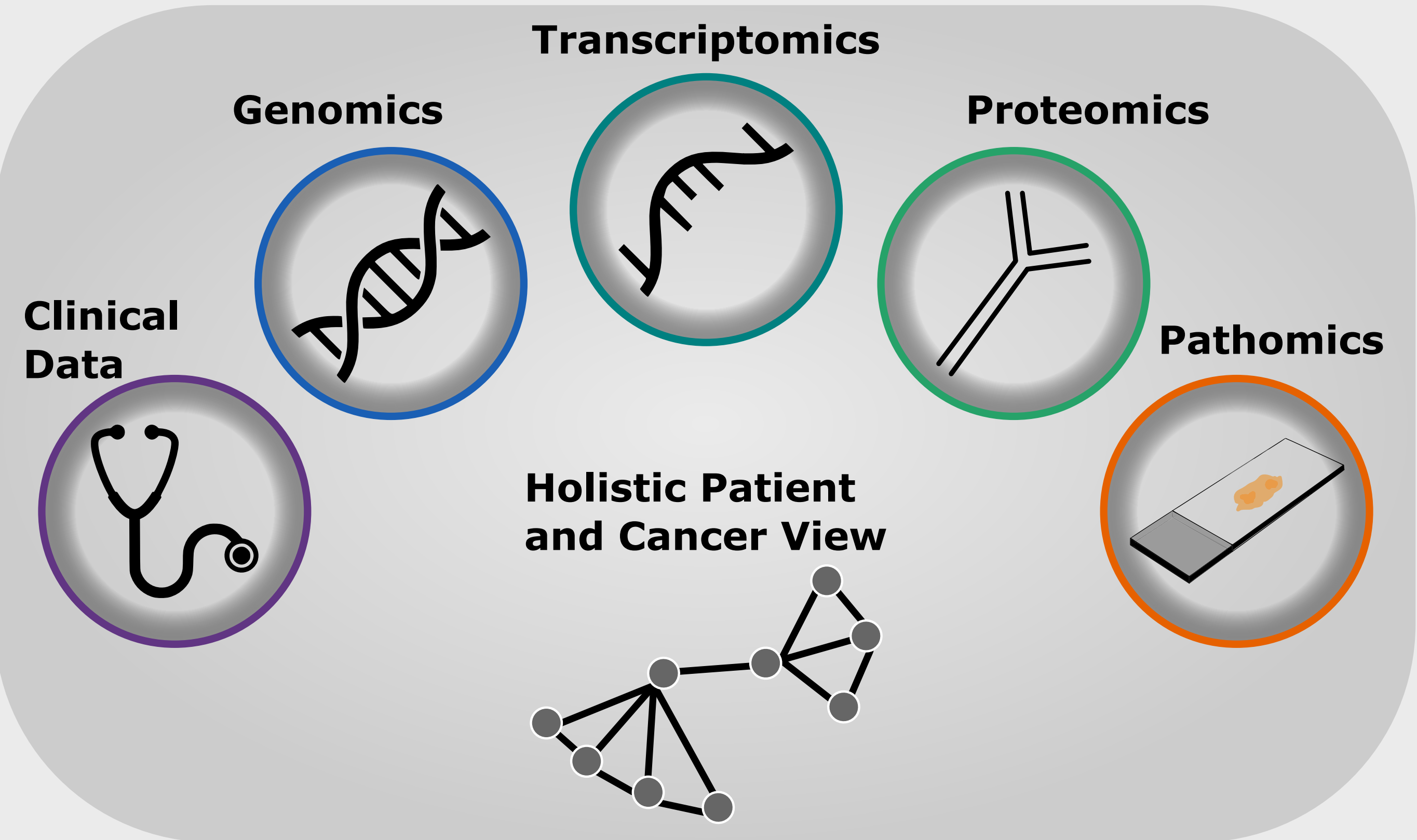The ROYAL MARSDEN
NHS Foundation Trust

Building a Platform for Integrative Discovery and Diagnostics in Cancer

Ferran Cardoso Rodriguez, Priya Lakshmi Narayanan, Constantino Carlos Reyes-Aldasoro, the Integrated Pathology Unit, Jacqueline James, Helen Colman, Maurice Loughrey, Tom Lund*, Manuel Salto-Tellez*
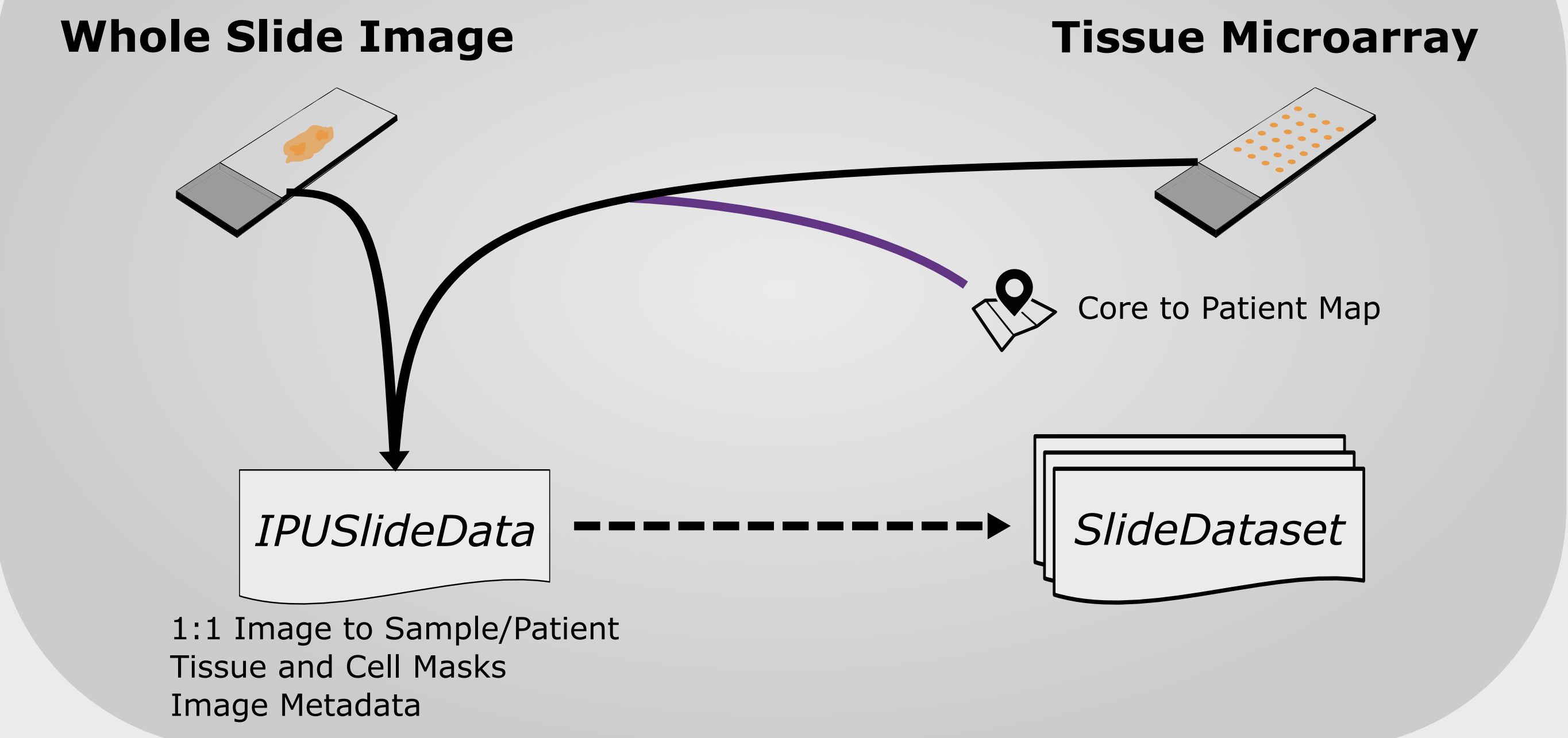
A Multimodal View of Patients and Disease

Transcriptomics
Genomics
Proteomics
Clinical Data
Pathomics
Holistic Patient and Cancer View

Digital pathology images are high-dimensional complex datasets routinely used in clinical and research settings. However, despite recent advances in other omic modalities, the field of pathomics remains relatively under-explored; with most of the readouts relying on low-throughput and semi-quantitative analysis by expert pathologists. Integrating images with clinical data and molecular omics modalities such as genomics, transcriptomics and proteomics is key towards constructing a multi-modal holistic view of patients and disease central to any Integrative Diagnostic and Discovery (IDD) platform.
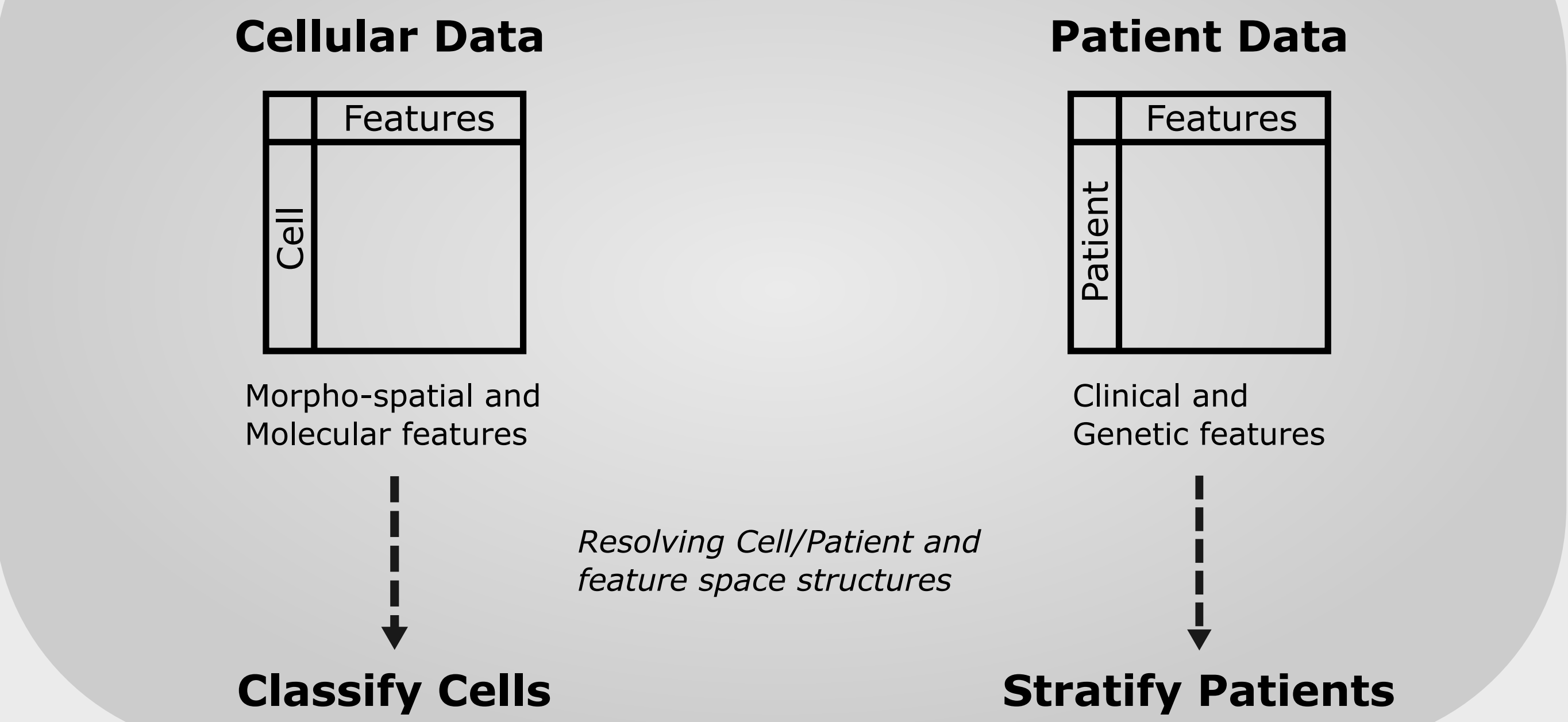
Building a Multimodal Image-centric Analysis Workflow

High-throughput processing of WSI and TMA cohorts

Whole Slide Image
Tissue Microarray
Core to Patient Map
IPUSlideData
SlideDataset

1:1 Image to Sample/Patient
Tissue and Cell Masks
Image Metadata

Generalisable approach where pathomic data are treated as a first-class modality, leveraging extended classes based on SlideData objects. Cohorts of WSI and TMA images can be analysed in a high-throughput unsupervised manner using algorithms for tissue core detection and mapping to relevant patient IDs. The resulting collection of sample or patient level objects can be grouped as SlideDatasets to undergo analysis and feature extraction in a local or distributed manner.
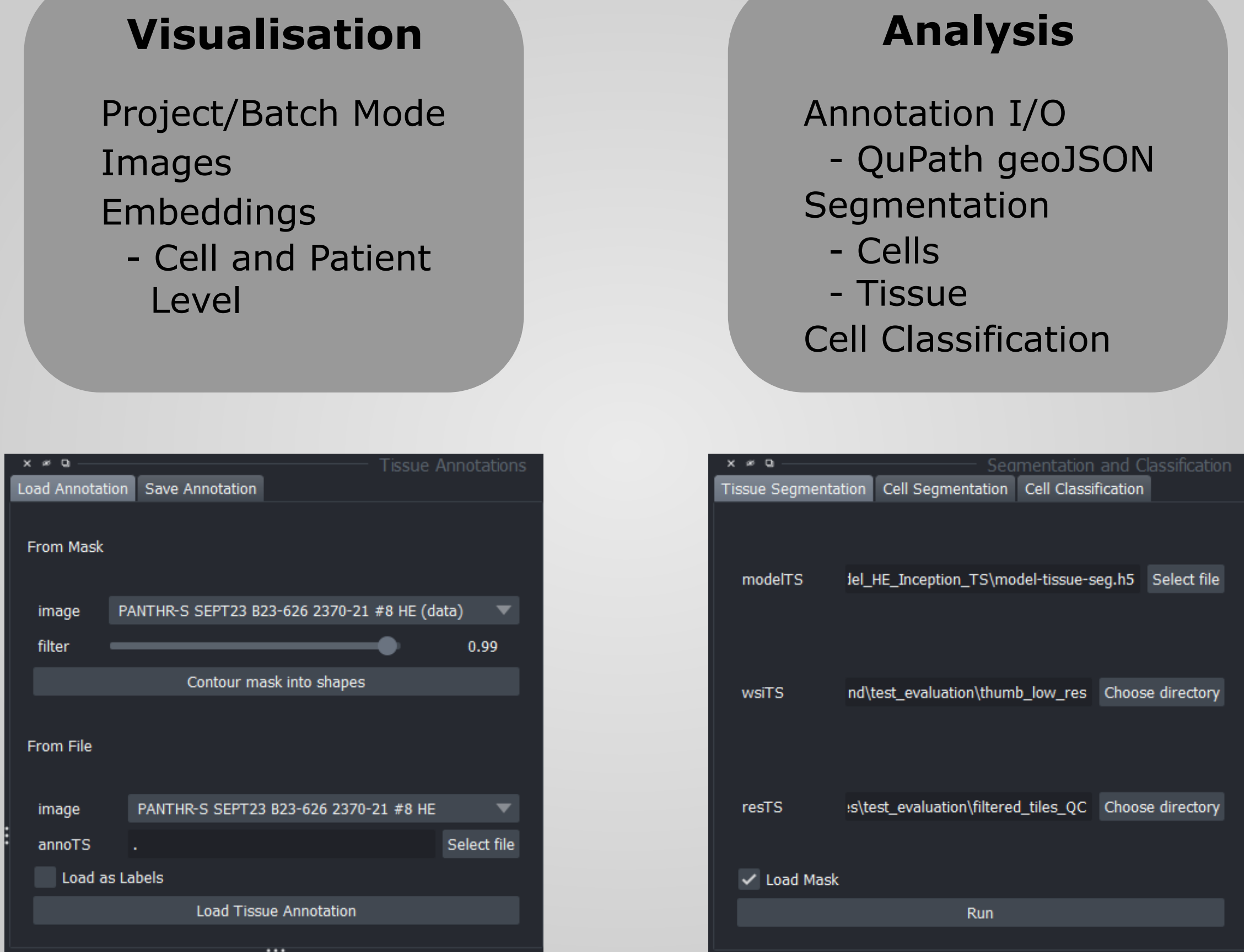
Common multiscale data structure

Cellular Data
Features
Cell
Morpho-spatial and Molecular features
Classify Cells

Patient Data
Features
Patient
Clinical and Genetic features
Stratify Patients

Resolving Cell/Patient and feature space structures

Revolving around next generation file formats and AnnData structures, the workflow captures a multiscale view at the level of cells and patients. Downstream steps aiming to resolve unit and feature structure in the different data spaces allow for cell classification and patient stratification tasks among others.
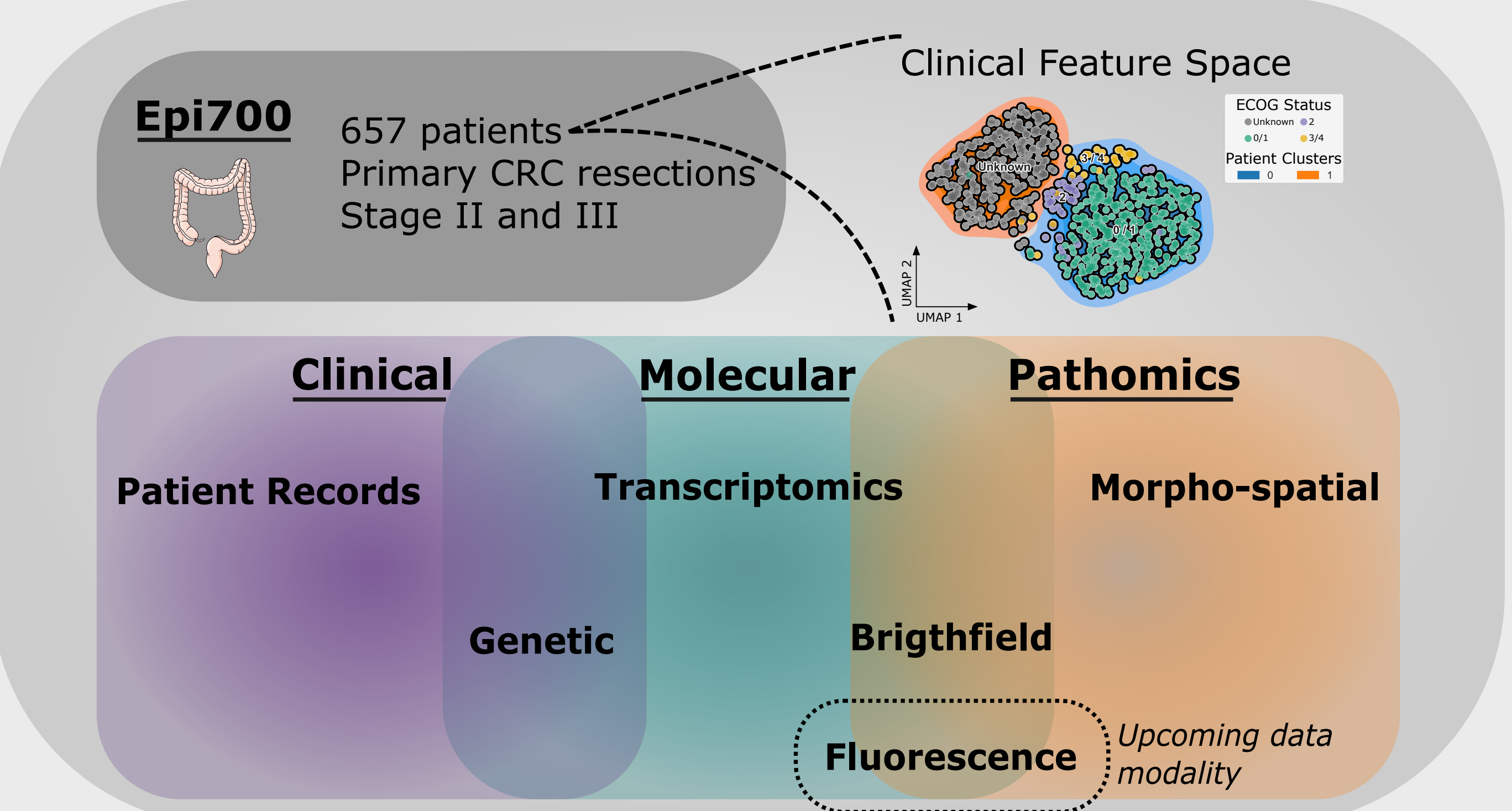
Bespoke Interface for Accessible Image Analysis

Custom Napari Front-End

Visualisation
Project/Batch Mode
Images
Embeddings
 - Cell and Patient Level

Analysis
Annotation I/O
 - QuPath geoJSON
Segmentation
 - Cells
 - Tissue
Cell Classification

Using Napari as a modular platform to visualise images and data, custom widgets implement back-end analysis steps in an accessible manner. Supporting the geoJSON format for QuPath interoperability, users can load, save and create annotations either manually or using deep learning models for tissue and cell segmentation and classification. Currently supporting off-the-shelf generalist models (e.g. cellpose) and custom in-house models, we plan to add human-in-the-loop training capabilities.

The Epi700 CRC Cohort as a Case Study

Epi700
657 patients
Primary CRC resections
Stage II and III

Clinical Feature Space
ECOG Status
Patient Clusters

Clinical — Patient Records
Molecular — Transcriptomics, Genetic
Pathomics — Morpho-spatial, Brigthfield
Fluorescence — *Upcoming data modality*

To showcase a real-world application of the IDD multimodal platform we chose the Epi700 CRC cohort of stage II and III patients from Northern Ireland. This cohort consists of clinical, molecular and digital pathology data layers, with the former used to generate a patient embedding coloured by ECOG status and Leiden clusters assignation. The addition of upcoming mIF images will help integrate the pathomics layer with the rest of the cohort to deploy our developing IDD platform.

Next Steps: Multi-level Heterogenous Graph

Molecular
Morpho-spatial
Cells
Patients
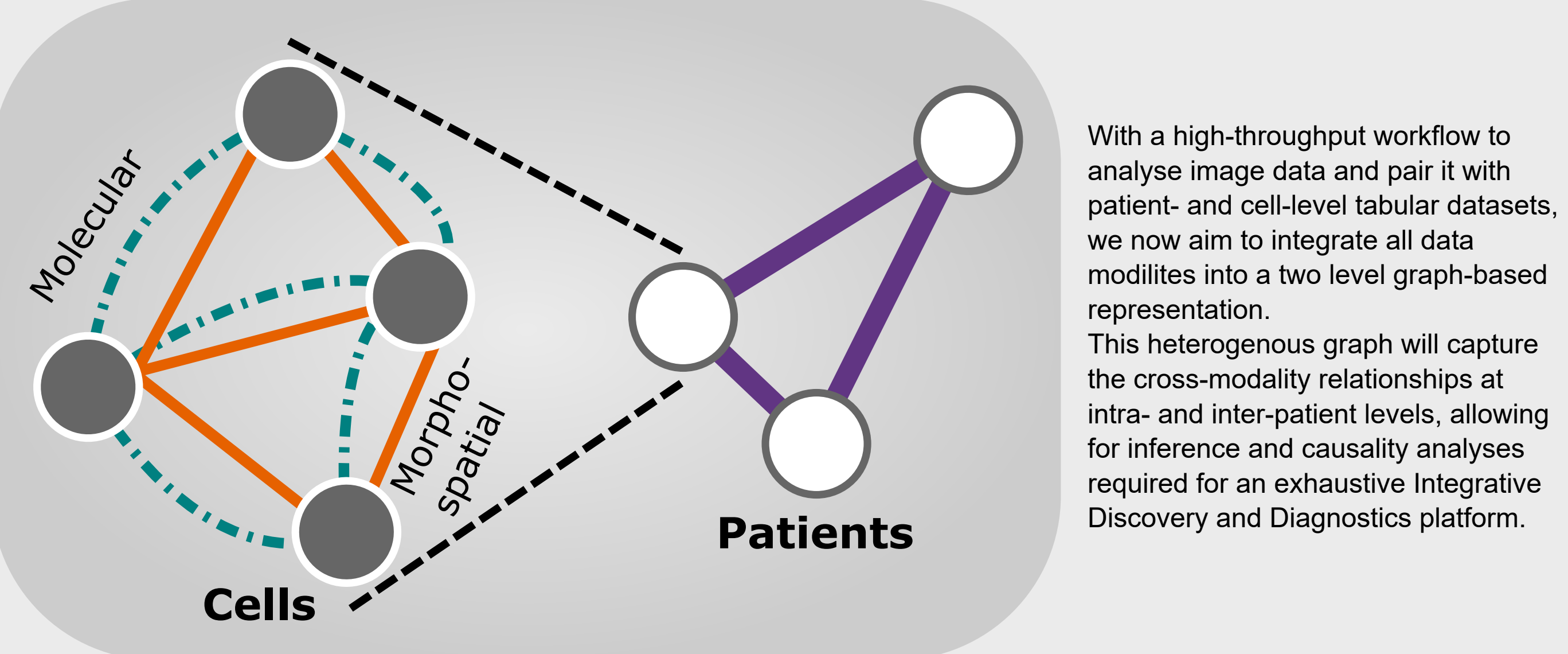
With a high-throughput workflow to analyse image data and pair it with patient- and cell-level tabular datasets, we now aim to integrate all data modilties into a two level graph-based representation.
This heterogenous graph will capture the cross-modality relationships at intra- and inter-patient levels, allowing for inference and causality analyses required for an exhaustive Integrative Discovery and Diagnostics platform.

References

Salto-Tellez, M., Maxwell, P. and Hamilton, P. Histopathology, 74: 372-376 (2019)
Bulow, R.D., Holscher, D.L., Costa, I.G. et al. npj Syst Biol Appl 9, 38 (2023)
Craig, S.G., Humphries, M.P., Alderdice, M. et al. Br J Cancer 123, 1280–1288 (2020)
Wagner SJ, Reisenbuchler D, West NP, et al. Cancer Cell, 41(9):1650-1661.e4 (2023)

Contact

Ferran Cardoso Rodriguez, PhD
Computational Senior Scientific Officer at the Integrated Pathology Unit
The Institute of Cancer Research & Royal Marsden Trust
T +442034376070 | E Ferran.Cardoso@icr.ac.uk