In-house and Commercial Platforms for Integrated Discovery and Diagnostics in Pathology

Ferran Cardoso Rodriguez [1], Priya Lakshmi Narayanan [1], Constantino Carlos Reyes-Aldasoro [2], Jacqueline A. James[3], Maurice B. Loughrey[3], Helen G. Coleman [3], Lucy Kilburn [1], Maggie Cheang [1], Tom Lund [1*], Manuel Salto-Tellez [1,3*]

[1]The Institute of Cancer Research & The Royal Marsden Hospital, UK

[2]Department of Computer Science, City, University of London, UK

[3]Queen's University Belfast, Northern Ireland

[*]These authors contributed equally

September 2, 2025

Abstract

Digital pathology (DP) is a key healthcare domain, generating large image-based datasets often assessed using semi-quantitative, low-throughput scoring by pathologists. With limited scalability and expanding cohort sizes, frameworks are needed to assemble high-dimensional pathomics datasets and enable discovery of data-driven biomarkers beyond traditional DP within both research and clinical settings.

We present two such frameworks: one developed in-house for a large colorectal cancer (CRC) Tissue Microarray (TMA) cohort, and another built on a commercial platform for data-rich clinical trials. The Pathomics Platform for Integrated Discovery and Diagnostics (POPIDD) is our modular in-house framework. POPIDD identifies, aligns and annotates TMA cores, links them with patient metadata, and supports pixel-, cellular- and regional-level analyses. Image and clinical data is standardised into tabular structures, with a graphical interface ensuring accessibility. The commercial framework applies Foundry, a generalist data platform, to retrospective and propestive studies with DP, molecular metrics and spatial transcriptomics; integrating multi-omics data over 3,000 patients and accruing multimodal data respectively.

Applied to the Epi700 CRC cohort (657 patients, >30,000 TMA cores), POPIDD reproduced known clinical and molecular associations, including greater immune infiltration in microsatellite instable and $BRAF^{MT}$ tumours. Within Foundry we built graph ontologies unifying data modalities and developed computational pipelines and accessible interfaces for both analysis and exploration. Incorporating a POPIDD image quantification module allowed us to quantify and compare cross-modality proliferation changes in response to therapy.

These case studies illustrate complementary strategies for advancing DP in multimodal settings. In-house platforms such as POPIDD provide flexibility and transparency but require technical upkeep, while commercial systems like Foundry facilitate clinical adoption through robust governance but at higher cost. Their parallel development is synergistic, as in-house or open-source tools can be incorporated into commercial frameworks.

Keywords— Digital Pathology, Pathomics, Computational Framework, Colorectal Cancer, Clinical Trials

References

[1] Craig, S.G. et al. Immune status is prognostic for poor survival in colorectal cancer patients and is associated with tumour hypoxia. Br J Cancer 123, 1280–1288 (2020). 10.1038/s41416-020-0985-5

[2] Ferran Cardoso. IntegratedPathologyUnit-ICR/pyhscore: Compute H-scores in Python. Zenodo (2024). 10.5281/zenodo.12531732