Building a Platform for Integrative Discovery and Diagnostics in Cancer

Ferran Cardoso Rodriguez [1], the *Integrated Pathology Unit*[1],
the *Precision Medicine Centre*[2], Tom Lund [1*], Manuel Salto-Tellez [1,2*]

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

[2]Centre for Cell Research and Cell Biology, Queen's University Belfast, Northern Ireland

*These authors contributed equally

April 30, 2024

Abstract

Digital pathology is a pivotal domain in global healthcare, where the images acquired represent 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 individual analysis by expert pathologists and of a qualitative or semi-quantitative nature. Moreover, the challenge of integrating image data with tabular and graph representations common in other omics modalities has isolated pathomics, posing a major issue for deriving diagnostics that often require of multi-modal analyses.

In this work, we propose a generalisable approach to integrate diverse data modalities with pathomics, inspired by practices and data structures used in single-cell omics applications. We have designed a standardised workflow to respect the secure data requirements of the healthcare domain, allowing for distributed computing and local or cloud deployment using third party solutions. The workflow we have developed leverages next-generation file formats and `AnnData` structures, resulting in a comprehensive platform to perform various tasks; from electronic health record and clinical data analysis with `ehrapy`, to image analysis and human-in-the-loop mode training using custom *Napari* plugins. When applied to the Epi700 colorectal cancer cohort we show how this in-development platform recapitulates observations from both clinical and imaging data spaces, and we also explore the potential of leveraging heterogeneous graphs to capture the representations of different modalities and their relations.

By building user-friendly graphical solutions around sections of the platform that most benefit from expert pathologist input, we have built an accessible and explainable platform to deploy machine learning models. The integration of pathomics with wider omic modalities allows for the construction of a holistic view with the potential to discover novel insights into the molecular mechanisms of colorectal cancer and to inform downstream integrated diagnostics.

Keywords— Digital Pathology, Pathomics, Multiomics Analysis, Integrative Discovery and Diagnostic

References

[1] Salto-Tellez, M., Maxwell, P. and Hamilton, P. Artificial intelligence—the third revolution in pathology. Histopathology, 74: 372-376 (2019). https://doi.org/10.1111/his.13760

[2] Bülow, R.D., Hölscher, D.L., Costa, I.G. et al. Extending the landscape of omics technologies by pathomics. npj Syst Biol Appl 9, 38 (2023). https://doi.org/10.1038/s41540-023-00301-9

[3] Craig, S.G., Humphries, M.P., Alderdice, M. 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). https://doi.org/10.1038/s41416-020-0985-5

[4] Wagner SJ, Reisenbüchler D, West NP, et al. Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study. Cancer Cell, 41(9):1650-1661.e4 (2023). https://doi.org/10.1016/j.ccell.2023.08.002