PathLAKE
Computational Pathology Excellence

Prof. Manuel Salto-Tellez

Contents

"One of PathLAKE's important goals is to bring new AI tools to routine tissue diagnostics. This is likely to have an influential legacy in years to come.

The diagnostic Digital Pathology exemplars created by each of the PathLAKE NHS and Academic partners involve different areas of pathology and different diagnostic purposes. Coventry/Warwick aim to make colorectal cancer screening faster and more affordable; Nottingham/Warwick's goal is to deliver standard of care in breast cancer diagnostics in a fully objective, reproducible manner; Oxford is transforming the way we approach prostate cancer at many levels of the diagnostic pipeline; finally, Belfast aims to deliver new predictive markers for colorectal cancer in the space of immuno-oncology.

Thousands of clinical, digital tissue images with clinical metadata are used to teach, validate and verify these new tools through a close collaboration between computer scientists and pathologists.

All the algorithms are in the late stages of development, meeting all initial targets of PathLAKE, and are now ready to be considered for clinical adoption with industry partners."
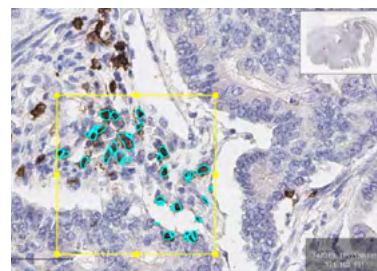
PathLAKE Exemplar Projects

The PathLAKE exemplar projects are developing AI tools with different purposes and tackling different areas of disease (including breast, prostate and colorectal cancer). We have made good progress, creating AI tools to assist Pathologists with making a diagnosis or to help predict outcomes (or the best course of treatment) for patients.

Belfast

The AI algorithms developed by the team at the Precision Medicine Centre of Excellence at Queen's University Belfast are new predictive tools to improve patient stratification in colorectal cancer. To do so, three AI algorithms have been created to help identify and quantify the number of three specific types of immune cells present within the colorectal cancer tissue, based on prior published experience **[PMID: 32684627]**. This is complemented with analysis using other sophisticated tools **[PMID: 32671911]** to accurately establish the ground truth of this test.

Over the past three years, we have scanned thousands of pathology slides and have created over 300,000 digital annotations on colorectal cancer samples previously analysed by chromogenic immunohistochemistry to help identify the immune cell markers of interest. The three algorithms are now being completed with some final testing using digital pathology images that have been shared with us by our PathLAKE NHS and academic partners. This helps to strengthen the ability of the algorithms to identify the immune cell content in images of tissue samples that may have some variation in appearance. These differences may be due to different processes used within the labs for creating the tissue slides, the way in which the slides are stained by different automated staining platforms or the different types of images created by different digital slide scanners. Exposing the algorithms to these differences in appearance helps to replicate how images are created within the clinical setting and means that the algorithms are better prepared for use in hospital pathology labs.

The best clinical performance is dictated by the combination of the three algorithms to create a single AI tool. An online predictor tool based on our experience of almost 2,000 samples analysed to date is currently under development. Our work will serve to seek regulatory approval with our industry partner for use of the tool as part of routine clinical practice within hospital labs in the NHS and further afield, helping pathologists bring real benefits to patients.

BRaCE

The Breast Pathology Research Group at University of Nottingham and computer scientists at University of Warwick's Tissue Image Analytics (TIA) Centre are working together on the PathLAKE Breast Cancer Exemplar Project - BRaCE.
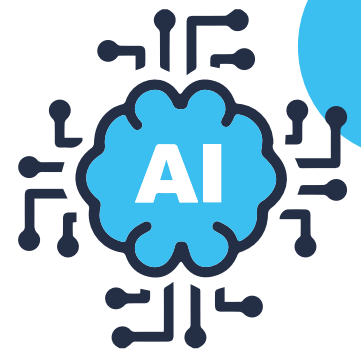
Using the vast archive of breast cancer tissue at Nottingham University Hospitals Trust, the aim is to create an algorithm that will classify the indeterminate risk group of breast cancer patients into precise treatment groups, depending on whether they will or will not benefit from chemotherapy treatment.

The project uses tissue and data from a large well-characterised invasive breast cancer cohort (2400 cases), with long term follow-up data and treatment information, to develop the prognostic algorithm, which includes a total of 8498 images.

The algorithm is being developed jointly by the team of pathologists at Nottingham, led by Professor Emad Rakha, and computer scientists at TIA, led by Professor Nasir Rajpoot. It is being expanded using different variables in breast cancer, for example by introducing tumour grade, shape and pattern of the malignant cells, the rate of tumour cell division and growth and the relation between the adjacent and surrounding cells to the cancer cells.

The BRaCE project incorporates seven sub-projects which are all contributing to the development of the final prognostic tool for breast cancer. The analysis is showing very promising and significant results. The final output is due soon.

CoBi

The PathLAKE Large Bowel Biopsy Screening Tool, co-developed by the PathLAKE teams at UHCW (led by Professor David Snead) and Warwick (led by Professor Nasir Rajpoot), is an algorithm that assesses digital images of large bowel biopsy slides ahead of the pathologist.

Patients who are suffering from symptoms of large bowel disease are most commonly investigated by colonoscopy, a flexible telescope used to look at the inner lining of the bowel. The clinician or nurse doing the examination will take small tissue samples, called biopsies, of any suspicious areas they find during the colonoscopy.

The biopsy samples are processed in the pathology laboratory where doctors trained in examining tissue (pathologists) look at them under the microscope. The process of examining the biopsies is very labour intensive and involves many hours of pathologist time to examine the samples under the microscope, first to find abnormal areas and then to decide what the disease is. A third of the samples taken are normal, but even these samples take many hours of pathologist time to report.
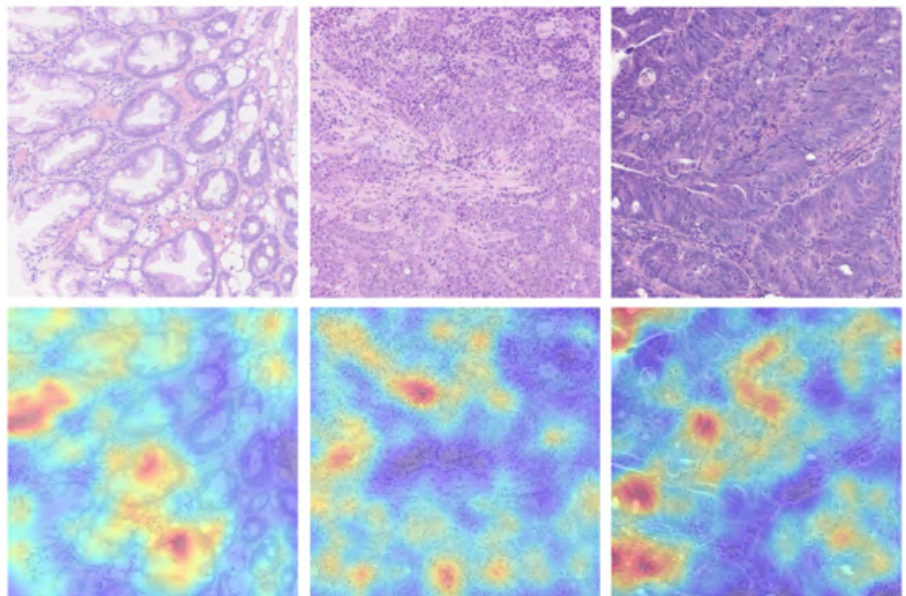
The biopsy is critical to finding disease and to understand precisely what the disease is (most often, tumour or inflammatory disease), how severe it is and how it should be treated. In some quite common disorders of the large bowel, the lining looks normal but contains microscopic abnormalities that can only be seen by the pathologist, the so-called microscopic colitis. Therefore, even for normal colonoscopies, biopsies are taken to check for microscopic colitis.

CoBi is an algorithm that assesses digital images of large bowel biopsy slides ahead of the pathologist.

The tool decides if the sample is normal and, if so, issues an automated report to the clinical team. If the sample is abnormal, the tool decides what the disease process is, where in the biopsy the abnormality is and places the case in an urgent or non-urgent basket.

As a result, the algorithm will save the time pathologists spend looking at normal biopsies so it could be used to examine the cases which contain disease and so we can be sure all cases which require urgent attention are examined first. All patients will benefit from this innovation. Patients without disease will get their normal biopsies reported earlier, and patients with disease will have their biopsies seen by the pathologist quicker. Pathologists will benefit by having the algorithm help identify areas of abnormality which will mean fewer errors through missing small areas of abnormality by oversight.

The tool is going through the proof of principle trials at the moment which will be completed by September. Our interim results are quite promising, with specificity of 80% at 97% sensitivity. We are now working very hard to move to the next stage, which will be a multi-site validation study across 10 sites in the UK which (if funded) we aim to start later this year. Efficacy and safety data from this study should enable the tool to enter practice in 2025.

> " **PathLAKE announces its first pathology robot.** "

Oxford

The PathLAKE team in Oxford has been working on several different projects as part of their "exemplar" workstream.

The work encompasses different aspects of digitization of histopathology and processes including adding images to the PathLAKE datalake, assessing image quality of slides and automating lab processes.

Staff in Oxford have curated 30,000 cases for the datalake. In another project, more than 11,000 slides have been retrieved from an off-site storage facility (a disused Welsh mine believe it or not!), curated, barcoded, cleaned and scanned. These types of large data sets (big data) can be used to find patterns of disease or identify risk factors for certain diseases and lead to better outcomes for patients. This particular library of cases is a hugely valuable research resource and will enable large scale projects looking at various aspects of prostate cancer.

Other work in Oxford has focused on different technical and mechanistic aspects, for example, one of the projects has led to the design and validation of an Automatic IHC (immunohistochemistry) Request Tool (**PMID: 34017063**).
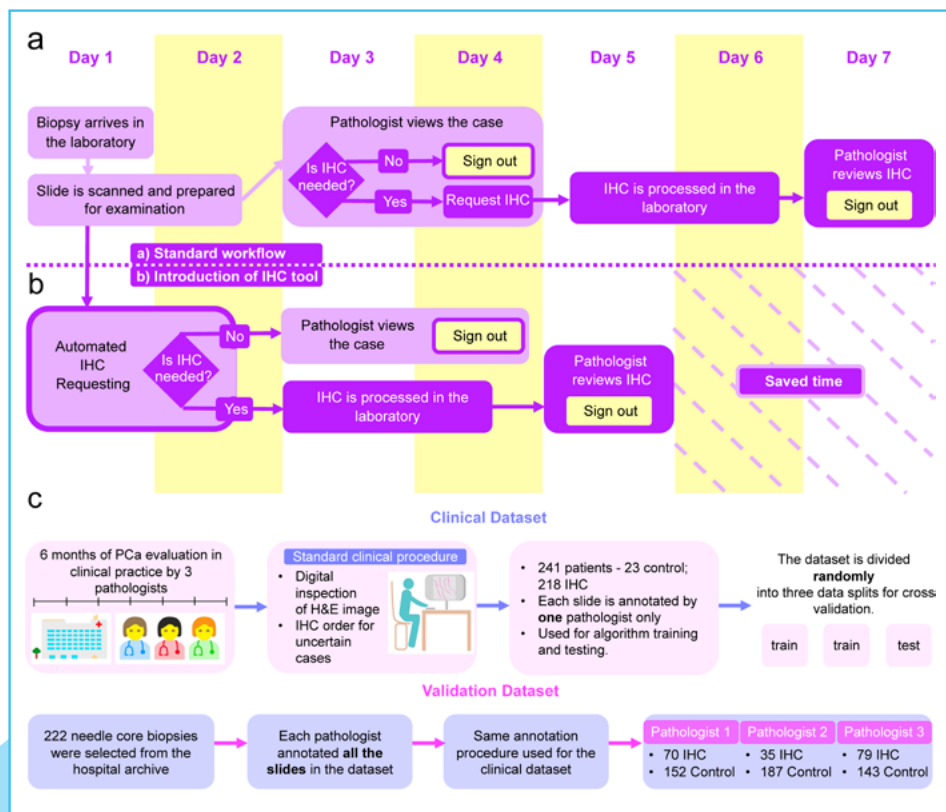
Pathologists normally review a case by looking at H&E (haematoxylin and eosin) slides, which is where prostate tissue is stained to show different tissue patterns indicative of cancerous or benign tissue. However, sometimes further information is required in order to give a diagnosis and that comes in the form of immunohistochemistry staining. Requesting this is a time consuming routine clinical task for the pathologist and means reviewing the case in order to request the IHC and then re-reviewing once that is available. The new tool uses AI (Artificial Intelligence) to automate the request for IHC for prostate biopsy cases without a pathologist reviewing the case first.

This means that the pathologist only reviews the case when all the necessary information is gathered together, enabling them to get results to patients and the treating clinicians in a quicker time frame. This time saving will also allow pathologists to review more cases, again, for the benefit of patients.

Our recent publication in Scientific Reports (Haghighat M et al) is one successful example of our pathology/engineering science collaboration and describes automated quality assessment of histology cohorts, unlocking their potential. **(PMID: 35322056)**

In addition to work on automation of workflows, the Oxford team have an ongoing package of work using AI to evaluate prostate cancer in novel ways and we should be able to share some of this work very shortly.

Events and News

Warwick Seminar Series

The TIA Centre Seminar Series welcomes external speakers to give presentations on Computational Pathology and aims to stimulate thought provoking discussions among participants. So far, we have had ten speakers from various countries, who have discussed their recent work in the field. Recorded sessions are available on YouTube and further information on upcoming seminars can be found on our webpage.

Click **here** for the YouTube channel and the **webpage**.

Visit to Belfast in November 2021

In November, Deborah Griggs and Rachel Flowers (PathLAKE Programme Manager and Project Officer) were delighted to visit Queen's University Belfast to meet the PathLAKE team at the Precision Medicine Centre of Excellence. After many virtual meetings over the last two years, it was great to meet face to face again at long last. Perry Maxwell gave us a guided tour of the Precision Medicine Centre labs and facilities, and Claire Lewis showed us around the Northern Ireland Biobank and answered our many questions.

Professor Manuel Salto-Tellez joined us for lunch and, in the afternoon, we met with Dominique French (Belfast Project Manager) to do a deep dive on the Work Package 3 deliverables and milestones and better understand the work lead by the Belfast team. Thank you to Prof Salto-Tellez and his team for making us so welcome, and for a very useful and informative day.

Precision Medicine Centre, Belfast

PathLAKE Showcase Conference – Live: 8 June 2022, Royal College of Pathologists, London

The PathLAKE Showcase Conference will bring together funders, academics and commercial partners to demonstrate its achievements and impact.

The PathLAKE project draws to a close in September 2022. Our showcase conference will demonstrate the significant achievements and impact of the project exemplars, before moving on to discuss data and specifically the challenges and opportunities in developing data repository systems and the ethics surrounding data sharing.

Nottingham Breast Pathology Masterclass – Online: 23 March 2022

The first Nottingham Breast Pathology Masterclass of 2022 gave over 330 delegates from across the world access to leading experts to discuss challenges in diagnostic breast pathology. The programme enabled delegates to enhance their knowledge, skills and expertise covering a range of different topics. The course leader, Professor Emad Rakha, was joined by a distinguished faculty: Professor Ian Ellis; Professor Puay Hoon Tan; Professor Maria Pia Foschini; Professor Abeer Shaaban; and Mr Hazem Khout. The Q&A sessions were moderated by Dr Islam Abdelaziz.

Masterclass – Live: 17 June 2022

Nottingham Breast Pathology Masterclass - Live takes place at the Crowne Plaza Hotel, Nottingham.

Cases will be presented in workshops with live microscopy and interactive discussion by our distinguished faculty consisting of Professor Ian Ellis, Professor Abeer Shaaban and Dr Elena Provenzano, along with Course Director, Professor Emad Rakha.

The Masterclass is ideal for practising pathologists, pathology fellows and trainees. Find out more **here**.

Publications

In each edition, we shine a spotlight on one of the most recent PathLAKE publications and interview some of the authors involved in creating these papers.

'Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study'

Mohsin Bilal is a Senior Research Fellow (Data Scientist) in the PathLAKE centre of excellence on AI in pathology at the University of Warwick.

To read the full paper in Lancet Digital Health, click this link: PMID: 34686474

Q1. Which molecular pathways of colorectal cancer can be determined computationally using routine histology slides in your paper?
Colorectal cancer is a heterogenous disease because of three major molecular pathways of its development: microsatellite instability, chromosomal instability, and CpG island methylator phenotype. We have shown that our algorithm can predict the status of all three molecular pathways, as well as tumours with high mutation density and key genetic mutations (BRAF, KRAS and TP53), of colorectal cancer by analysing the whole slide images of routine histology slides.

Q2. How does your algorithm predict the status of molecular pathways from routine histology images?
Our algorithm analyses raw pixel data of whole slide images to find image regions with differentiating patterns related to the status of molecular pathways. We first train a deep neural network model using a novel IDARS training strategy on a retrospective cohort, which contains whole slide images belonging to two different classes, each representing a different molecular pathway e.g. microsatellite unstable vs microsatellite stable tumours. Testing of the trained model on internal and external test sets determines the predictability of the model. The area under the receiver operating curve (AUROC) of above 0.8 on the test set for a given prediction problem defines the

success of the model. For example, our model predicts microsatellite instability with AUROC = 0.86, which is quite promising. This shows a great potential of deep learning to find differential patterns from the available cohorts of whole slide images with slide-level labels without requiring any regional or cellular annotations. However, this way of determining status of a molecular pathway or genetic mutation is correlative but cannot be considered causal. A major side benefit of our method is that it enables identification of differential visual and sub-visual patterns within each group. These patterns may or may not be truly associated with the molecular label being predicted. However, we have shown with further analysis and visualization of the discovered patterns to be statistically significant in verifying the prediction of molecular pathways. Differential cellular composition of highly predictive regions for the given molecular label also offers a genotype-phenotype mapping with an opportunity to add explainability to the model's predictions and to create new knowledge and novel insights about the disease heterogeneity.

Q3. What is the impact of determining molecular pathways computationally?
The key to diagnosing and treating colorectal cancers efficiently lies in understanding the pathways and key mutations involved in their development. Computational methods can improve clinical practice by efficiently determining the status of molecular pathways and key mutations. This can save time and money, and decrease the rate of genetic testing.

Q4: What is the current state of research and its adoption in the clinical practice?
The results of our research so far are quite promising. However, further research is needed to benchmark the baseline performance on independent and large-scale multi-centric cohorts before clinical adoption.

Please follow the link here to find a list of our most recent PathLAKE publications

Contact Us
☎ +44 (0) 2476 968 582
✉ pathlake@uhcw.nhs.uk
🖥 www.pathlake.org

PathLAKE Office
Clinical Sciences Research Laboratories
University Hospitals Coventry and Warwickshire NHS Trust
Clifford Bridge Road, Coventry CV2 2DX

Follow Us
🐦 @PathLAKE_CoE