Name: Fidelma Kaar
Year Bursary awarded: 2024
Medical School: University College Cork
Topic of Elective: Digital pathology
Department where elective was completed: Department of Histopathology, Galway University Hospital and Department of Histopathology, Cork University Hospital.
Supervisors: Professor Louise Burke (MD FRCPath FRCPI FFPath (RCPI)) and Professor Sean Hynes (BSc PhD MB BCH BAO FRCPath(UK) FFPATH (RCPI)).

Summary of project:
There were two parts to this study, a breast cancer study (part A) and a lung cancer study (part B)

Part A:
Digital pathology has seen major advancements in recent years. Tumour specimens can be digitised and analysed using machine learning and Artificial intelligence (A.I.) techniques. By analysing the morphology of tumour cells in digitized pathology images, machine learning models can be trained to identify patterns associated with specific tumour genotypes including specific mutations and genetic markers. These approaches may aid in diagnosis and may have implications for personalized medicine and treatment planning by identifying specific genetic characteristics of tumours that may influence response to therapy. Specific genetic tests used in the diagnostic process for cancer treatment are expensive and not available in many countries. Such tests include Oncotype DX, a 21 gene assay used in breast cancer treatment planning and prognosis. We predicted the ability to use of machine learning algorithms to predict a 'high' or 'low' Oncotype score based on the clinical cut off of 16 in breast cancer specimens.

Aim: The aim of this study was to use machine learning, algorithmic techniques to predict genotype from phenotype.

Methods:
In the breast cancer arm of the study 30 slides from patients with advanced stage breast cancer which was previously subject to Oncotype assessment were digitised to create whole slide images (WSIs) and analysed using QuPath image analysis software. 28 WSIs were used; 10 x >16, 8 x <16 and 10 'unseen' oncotype scored specimens. In experiment 1 classifiers were trained on the >16 and <16 scored WSIs. Classifiers were applied to the unseen WSIs. In experiment 2 classifiers were created using three groups, a 'High'(20+), 'Intermediate and a 'Low'(<11) oncotype group and applied across all WSIs.

Results:
There is no significant difference between >16 and the <16 WSIs (p= 0.9691 and 0.8589 respectively). Normalisation of data using tumour percentage does not affect this (p=0.891 and 0.6205 respectively). However, we noticed less tumour cell detection in the WSIs with scores which fell close to or at the clinical cut off. Similarly in experiment 2 less tumour cells were identified in the intermediate oncotype group by all three classifiers. When the 'High' classifier was applied to all WSI groups there were significantly less tumour cell detections in the intermediate group compared to the high group and low groups (p=0.0181 and 0.0376). We propose a difference in morphology as recognised by the algorithmic models in the intermediate group compared to the high or low groups which was recognised in both experiments. We suggest

that an algorithmic, machine learning approach is not sufficient to explore this difference, and future work may involve the use of deep learning AI methods.

Part B:
Intra-tumoral heterogeneity describes the phenomenon that cells within a tumour will exhibit different phenotypes and morphologies. Such variability must be considered throughout the diagnostic process in order for accurate diagnosis of the tumour type. Lung adenocarcinoma often contains extensive intra-tumoral morphological variability on histological analysis. Intra-tumoral heterogeneity also provides an explanation for the existence of sub-clonal diversity within a single tumour. Due to this sub-clonal diversity which occurs throughout the evolution of a cancer, lung cancer often contains genetic variants which can become treatment resistant and drive cancer progression. This heterogeneity has therefore posed enormous challenges when it comes to therapeutics for lung cancer.

Follow up of lung cancer patients can be done in the form of a blood draw called a liquid biopsy which may contain circulating cell-free tumour DNA (ctDNA) or circulating tumour cells (CTCs). The DNA within the liquid biopsy is extracted and subject to DNA sequencing (Next generation Sequencing) to identify clones that have become resistant to therapeutics. In approximately sixty percent of patients this resistance develops as a result of selection for clones that harbour a secondary mutation in EGFR called T790M. Third generation tyrosine kinase Inhibitor (TKI) drugs such as rociletinib can successfully target both the activating EGFR mutation and the T790M mutation.

In the current study we are interested in matching an area within the primary lung tumour using digital pathology techniques to with a resistant clone in order to potentially develop a digital biomarker that could possibly herald tumour progression in lung cancer patients. Artificial intelligent (AI) algorithms such as machine learning and deep learning are being employed for detection and classification in digital pathological images. Previous studies have investigated the concept of training AI algorithms to match morphological features to a specific mutation. One such study by Coudray et al explored this concept in non-small cell lung cancer (1). This study trained an AI model to successfully predict six commonly mutated genes in lung adenocarcinoma.

Aim:
The aim of this study was to develop virtual biomarkers that could potentially identify treatment resistant clones using tumour morphology and digital pathology.

Methods:
This study was carried out in a fully informed consented manner using a consent form and information leaflet for patients who were attending CUH Medical Oncology Department for lung cancer follow up with liquid biopsy. Consented patients gave an extra blood draw at this appointment in order to provide a validation of liquid biopsy in CUH using Next Generation Sequencing of tumour DNA in their blood. When results from this liquid biopsy showed the presence of a resistant clone we will then reviewed the histopathological reports of these patients (maximum of 20) in order to find a suitable primary lung cancer resection specimen from one of these patients. Sections from this patient's original primary lung resection were scanned to create whole slide digital images. These whole slide images (WSIs) were then be examined using digital pathology techniques on QuPath. Object classifiers can be created form the slide image

which employ machine learning techniques to train the programme to detect a specific cell type or morphology. Classifiers were created to identify morphologically distinct areas in the three whole slide images of the heterogenous primary tumour sections from a single patient. There were 4 chosen distinct areas; micropapillary, acinar, solid and mixed (micropapillary and acinar). These chosen areas were then micro dissected from the section and subject to single gene testing (Idylla PCR). We then compared both the sequencing results from the liquid biopsy and distinct morphological area within the primary tumour.

Results:
Using Single gene testing on the original biopsy detected a single EGFR L858R mutation within exon 21. However, liquid biopsy NGS detected a dual compound EGFR L833V/L858R mutation. This rare complex variant was confirmed within the resection specimen via NGS orthogonal testing. Identical clonal mutations for this variant were confirmed using NGS across multiple morphologies as detected by QuPath. However, a mixed morphology as detected by QuPath demonstrated significantly different allelic frequencies for this complex variant (15.5% vs 30.1%) compared to the other morphologies identified by QuPath which included acinar, solid and micropapillary.

This study demonstrates a relative stability of a driver compound mutation within the blood whilst on treatment. Furthermore, the use of open-source image analysis software could help in enriching/avoiding morphological areas to identifying specific mutational types and allelic frequencies which could impact downstream treatments and indicates that this is a promising avenue for further investigation.

What proportion of the project was histopathology?
Part A: 100% of the work was histopathology as 100% of the work was analysis of digital pathology slides.
Part B: My project is part of a greater study looking at the validation of liquid biopsy for lung cancer in Cork university hospital. My part of the study was 75% histopathology.

Has this experience changed your views about histopathology and research? If so, in what way?
This work has further enriched my knowledge in the area of digital histopathology. Machine learning and AI are revolutionizing diagnostic histopathology. This work has truly highlighted this for me and as a student has shown me that a small piece of work in the area of digital histopathology can contribute to a greater understanding of how the computer ‘sees’ the specimens and how we may harness this to our advantage as we move towards personalised medicine.

Have you presented this work anywhere? If so, where?
I have presented part of this work at the ISSP meeting, 2024.

References:
Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat Med. 2018 Oct 17;24(10):1559–67.