Spatial statistics and Hidden Markov Modelling to improve the efficiency of a novel diagnostic test for cancer using Single molecule imaging.

supervised by

Dr Hannah Mitchell and the STFC

This project is in the rapidly developing field of single molecule imaging, that seeks to gain biological insights on the scale of 0-200nm (below the diffraction limit of conventional light microscopy). Working closely with collaborators at STFC, world leaders in the development and exploitation of single molecule imaging techniques, this project seeks to facilitate the automatic exploitation of spatial-data derived from a single molecule imaging technique pioneered by the OCTOPUS group called Fluorescence Localisation Imaging with Photobleaching (FLImP) the subject of two Nature Comms papers [1,2]. This technique is used to measure the arrangement of protein clusters (oligomers) on the cell membrane. The different conformation these proteins adopt is known to play a role in the development of cancer. The ability to measure the population of shapes of these protein clusters adopt can be used as a novel diagnostic test for cancers such as small cell lung carcinomas.

As every patient's cancer is unique, this work is being undertaken with a view to developing novel diagnostic tools in order to rapidly fingerprint protein oligomerisation states in order to determine the most effective therapeutic strategy for each patient. Developing clinically translatable tools for such a personalised medicine approach requires full automation of the imaging technique in conjunction with fully tractable statistical techniques for the rapid analysis and exploitation of data generated, which these projects seek to support.

The FLImP single molecule imaging technique was pioneered by Professor Martin-Fernandez, Dr Rolfe and Dr Needham is capable of resolving molecular separations on the scale of 5 nanometers. FLImP has been used to investigate the oligomeric organisation of Epidermal Growth Factor Receptor (EGFR), which plays a central role in many human cancers. The OCTOPUS group at the STFC have been working to extend the FLImP technique to two-dimensions and translate this technology to the clinic for use as a tool for cancer diagnosis with collaborators at Kings College London. As part of this work Dr Davis & Dr Rolfe have been automating the FLImP acquisition and analysis technique to increase throughput. To date, the £1.3 Million FLImP microscope facility at STFC is generating 1 Terabyte of imaging data every 24 hours from which molecular separations are automatically determined and analysed. A key stage in the analysis process is rapidly determining which of the 1 million or so fluorophores extracted each day is suitable for FLImP analysis. This is achieved by labelling and segmenting 1-dimensional plots of integrated fluorophore intensity profiles over time containing two or more fluorophore photobleaching events. As the behaviour of each fluorophore can be described using a markov chain, an efficient Hidden Markov modelling approach could be used to maximise information extraction from these datasets.

A second arm of this project is concerned with improving the automation and exploitation of spatial statistical data generated by a second, complimentary, single molecule imaging microscopy technique called MINFLUX. MINFLUX is a recently developed single molecule imaging technique capable of resolving fluorophores in 3-dimensions to a resolution of 1-3nm

[3]. While not as mature as FLImP, this exciting technology has the potential to transform our understanding of biological systems at this spatial scale. Currently, analysis of MINFLUX data is extremely heuristic and labour intensive. For wide adoption of this technology, the development of automation and novel analysis tools are required. MINFLUX data is acquired in the form of Neyman-Scott point pattern processes, where daughter locations represent fluorophore localisations (of which there are many) and the unseen parent points represent the most likely position of the emitting fluorophores.

Project An opportunity has arisen for a statistically gifted PhD student to develop a novel Hidden Markov model to facilitate automatic labelling and segmentation of FLImP suitable fluorophore bleaching events from these integrated intensity profiles from simulated and real-world data. The objective of this project will be to develop the algorithms behind the hidden Markov model. The model will need to be robust to account for noise, partial information (two or more fluorophore bleaching events may happen at the same time) and any potential outliers within the data and will need to be computationally efficient to ensure it is practical to apply at volume. The model will also need to take into account the replicated experimental processes with the incorporation of random effects. Hidden Markov models such as the factorial hidden Markov model and the hidden semi Markov model will be investigated and the algorithms behind the models (Viterbi, Forward- backward) developed. The ultimate goals of this project are to maximise information extraction from the existing FLImP data archive (¿450 Terabytes of FLImP imaging data) and accelerate future FLImP data acquisition. This fits into my current research in terms of hidden Markov model development.

As part of this project we seek to develop novel techniques to rapidly assess the information content of sampled Neyman-Scott processes (essentially MINFLUX datasets as they are being acquired) to rapidly evaluate the likely quality of a region of interest. This is an important research question given the considerable time required to image a region of interest using MINFLUX, and the difficulty of making this distinction by eye. The main objective of this project will be to develop machine learning algorithms in conjunction with spatial point pattern analysis and taking into account fractal analysis. We are also seeking to reconstruct the most likely parent process that gave rise to a series of MINFLUX daughter localisations and from this parent process, reconstruct the most probable molecular structure(s) that gave rise to this process. This will involve the development of spatial statistics to understand the parent process within the Neyman-Scott point pattern and develop techniques to be able to uncover these as well as the molecular structure that gave rise to the process.

References:
[1] Nature Comms 9, Article number: 4325 (2018)
[2] Nature Comms 7, Article number: 13307 (2016)
[3] Nat Methods 17, 217–224 (2020)
[4] JORS 7, 1-13 (2020)
[5] SMTDA, 79, (2018)
[6] CASI, (2018)