Preliminary Protein Network Analysis in TIS Images of Colon Cancer

Nasir Rajpoot (Computer Science) & Mike Khan (Life Sciences)

A major goal in cancer research is to find biomarkers for early diagnosis — this is the ultimate aim of our project. Biomarkers can support earlier, more accurate, diagnosis and can be employed for disease and treatment monitoring. The ideal biomarkers can be found in readily accessible body fluids, such as blood and can thus be determined by using minimally or non-invasive techniques. For colorectal cancer (CRC), there is great interest in finding biomarkers in blood or via examination of stool samples or of cells captured from stool surface. Thus, one could avoid colonoscopy and biopsy, which is relatively costly and resource-intensive and in rare cases can be associated with life-threatening perforation. However, to date, successful use of tests for occult blood in stool samples notwithstanding, no sufficiently specific or sensitive tests are currently available. It has also been reported that single biomarkers may not meet these demands.

In this project, we will conduct a preliminary exploratory analysis for combinations of proteins that may reliably distinguish cancer from normal and potentially even provide information on prognosis and disease sub-classification. We will use image data generated by Mike Khan's research team (and possibly also data collected by a MOAC student in a linked experimental biology mini-project) using the Toponome Imaging System (TIS) multiplexed fluorescence microscope installed at the Wellesbourne campus. TIS allows the simultaneous analysis of multiple biomarkers and the potential to look at co-expression of markers in cancer cells, stroma etc, is likely to allow more accurate characterization.

One of the major challenges in this research is analysing the large amounts of data (several gigabytes per patient sample) generated by the TIS machine. As is often the case with fluorescence microscopy, the data may be corrupted with noise and other artifacts and the individual TIS channel images may not be aligned with each other due to various reasons. In another ongoing project funded by the Institute of Advanced Studies (IAS) at Warwick, we have developed sufficiently robust techniques for pre-processing of the TIS image data, thus facilitating the subsequent exploratory analysis to be carried out in this project.

An important feature of this mini-project will be a comparative analysis of several exploratory data analysis algorithms [1-3] on the TIS data generated from normal and cancerous specimens. This will involve looking for combinatorial molecular patterns which may enable one to differentiate between normal and cancerous cells and potentially allow one to characterise pre-cancerous stem cells.

A long-term goal of this research (a **PhD project** in itself) is to find protein complexes which may be enriched in local and distant metastatic sites and are, therefore, likely to represent functionally important complexes. Similarly, enrichment of stem cell signatures in advanced/metastatic lesions would strongly support the role of stem cells in driving malignant behaviour. This is an ideal PhD project as the technologies are established, limiting the risk to the project and yet there are ample opportunities for developing and refining the methods, as well as for answering several major questions in cancer biology and oncology.

Bibliography:

[1] van der Maaten, L. and G. Hinton, Visualizing data using t-SNE, Journal of Machine Learning Research, 2008. 9:p. 2579-2605.
[2] Frey, B. and D. Dueck, Clustering by passing messages between data points. Science, 2007. 315(5814): p. 972-976.
[3] Rajpoot, N. and M. Arif, Unsupervised Shape Clustering using Diffusion Maps. The Annals of the BMVA, 2008. 5: p. 1-15.