DESCRIPTION FOR THE GENERAL PUBLIC

The subject matter of the project should be classified as bioinformatics, and deals with developing new computational tools to aid in molecular imaging of cancer tissue. Bioinformatics is a branch of science that bases on the application of computer science and mathematical models to problems in molecular biology and medicine. Molecular imaging of tissues is a novel molecular biology experimental technique, which allows for a far more in-depth assessment of the spatial and molecular structure of tissues than any prior methods.

Molecular imaging involves a number of steps. Tissue specimens obtained from surgically extracted tumours are fixed, and a grid of measurement point locations is created on the sample's surface. Each of those points is subsequently hit hundreds of times by a laser beam. When the laser beam hits the tissue, some of the biological material becomes ionised and detached from the primary tissue, forming what is known as a "plume". The ionised biomolecules are led into the vacuum chamber of the mass spectroscopy unit and move across an electrostatic field towards an ion detector. The time between turning on the electrostatic field and the moment when the biomolecule hits the ion detector is measured, allowing us to "weigh" the biomolecule by using the fact that the acceleration of an ionised molecule in an electrostatic field is inversely proportional to its mass. The number of detector hits for each putative mass of a molecule can be combined into a graph known as a mass spectrum. This procedure is known as MALDI-IMS (Matrix-Assisted Laser Desorption Ionization – Imaging Mass Spectrometry).

Mass spectra obtained for the different points on the grid on the tissue are often used to construct maps (two-dimensional images) of the intensity of individual biomolecules across the tissue. Cutting edge MALDI-IMS spectrometers typically work with a 100-by-100 point grid, with the individual points being between 10 and 100 μm apart. This level of resolution allows interesting insight into the biology of cancer development. Mass spectra obtained for each of the grid's points include information on the intensities of hundreds of thousands of biomolecule masses. A single clinical trial can scan hundreds of tissue samples, leading to dataset volumes in the order of terabytes. Dealing with data of such scope is a great challenge to bioinformaticians, both in terms of basic operations such as transferring or storing the data as well as its efficient analysis given hardware limitations. Methodology for analysing such data has been rapidly developing over the past few years, but given the sheer volume of the datasets and their incredibly complex structure, the existing algorithms are quite limited in scope. Currently, tool development is focusing on the primary computational problems that need to be dealt with to aid molecular imaging, such as data compression, the identification of key molecular image traits, as well as the development of appropriate statistical testing for MALDI-IMS data. Ideally, such analysis tools should be presented in the form of integrated computational environments for ease of use.

The novelty of the proposal is the creation of a number of dedicated, efficient algorithms resolving a number of current MALDI-IMS data analysis shortcomings. Our prior work involved effective mathematical modelling of mass spectra. The application of Gaussian Mixture Models to the research problem at hand will allow us to clearly identify the primary traits of molecular images, as well as compress the data 20-30-fold. We plan to tailor the methodology to the nature of the data, develop big data clustering and statistical analysis based on the identified molecular image traits, and wrap up the resulting algorithms into a single computational environment for data analysis and visualisation. The algorithms will be validated on a number of actual tumours datasets, featuring the analyses of head and neck, skin, breast, thyroid and prostate cancer.

The support of MALDI-IMS through the development of an appropriate computational framework will also aid cancer diagnostics and therapy planning. This is of great significance to the research planned within the project. The development of computational approaches and dedicated mathematical modelling for the data will allow for in-depth comparisons between cancerous and healthy tissue, as well as detailed analysis of inhomogeneous tumour regions and their corresponding metabolite profiles. This will lead to increased insight into tumour development, allowing for improved therapy planning and reducing the social costs of cancer.