Project title: Developing dynamic latent process methodology for high dimensional biomarker data with correlated latent biological processes

Supervisor(s): Brian Tom and Robert Goudie

Abstract:
In the area of precision medicine, the understanding of the biological mechanisms underlying disease and defining a new molecular taxonomy of disease are of particular interest. Heterogeneity in disease phenotypes can be described either by discrete disease subtypes or by a continuous disease spectrum arising from the relative contribution of a number of drivers or latent factors at play at any point in time. For example, under the former perspective, there may exist subpopulations of patients with different disease progression due to having different biomarker profiles. Under the latter viewpoint, the drivers may represent a modest number of important pathological mechanisms (or pathways). For example, they may relate to immune response, dysregulation and inflammation in auto-immune diseases.

When adopting the latter perspective, Bayesian (dynamic) latent factor analysis methodology may be particularly appropriate. In this PhD project, we plan to extend this methodology to the situation where individuals with disease have intermittent follow-up visits and high dimensional biomarker measurements taken longitudinally. Specifically, it will build on earlier work done by Carvalho et al. (2008) who implemented a high-dimensional sparse factor model, and Chen et al. (2011) in which a modelling approach based on splines was proposed. Particular challenges include the determination of the number of unknown latent drivers, how best to model these underlying latent processes when correlated and how best to introduce sparsity constraints on the factor loading matrix so as to make the methodology scalable to the high-dimensional setting. Moreover, there are computational challenges that may arise due to poor mixing and weak identifiability. Use of prior biological knowledge and access to control/reference data may help alleviate some of these challenges. How best to incorporate these pieces of information into the modelling framework is an additional open problem.

This project would suit a student interested in developing, implementing and applying novel statistical methodology, with an interest in Bayesian methodology and high-dimensional data. The project will require creativity from the PhD student to develop, critically-appraise and implement their own research ideas in discussion with their supervisor.

1. Carvalho et al. (2008) High-dimensional sparse factor modelling: Applications in gene expression genomics. JASA, 103; 1438-1456
2. Chen et al. (2011). Predicting viral infection from high-dimensional biomarker trajectories. JASA, 106; 1259-1279.

Start Date: Michaelmas Term 2021 (all applicants)

All application queries regarding eligibility should be directed to phdstudy@mrc-bsu.cam.ac.uk

How to Apply: Applications should be made on-line via www.graduate.study.cam.ac.uk/applicant-portal selecting course details MDBI22 PhD in Biostatistics

Deadline for applications to be considered in the initial round of recruitment: 7th January 2021