Construction of multi layer gene regulatory networks to understand the human immune system.

**Supervisor:- Chris Wallace**

The genes that control how our bodies function do not operate in isolation but in cooperation with each other in complex networks. Indeed some have considered these networks might be so extensive, that they have proposed an omnigenic model in which (nearly) every gene interacts with (nearly) every other gene indirectly. However, even in this model, there is considered to be a “core” set of genes whose function directly impacts on human phenotypes such as disease.

This project will explore and extend methods for mapping such core sets by inferring gene regulatory networks using human gene expression data. In particular, the focus will be (1) on methods that move beyond pairwise correlation of gene expression to include genetic data which will be used to help assign direction to edges, and (2) methods for multilayer networks that allow for networks to vary between cell types and for interactions to exist between cell types. Regularized regression will be used to reduce overfitting when the presence or absence of network edges and the magnitude of their influence must be learnt simultaneously from the same data. Novel work will examine construction of variable penalty parameters that encode known external functional information about gene pairs (eg location of gene product within or outside the cell, established protein-protein interactions).

The intended application is to learn the regulatory networks that control the healthy human immune system, and to explore how altering expression of one or several genes (via genetic mutation) can change its function. Ultimately, this will help us understand how dysregulation of the immune system contributes to immune mediated disease such as rheumatoid arthritis. However, the statistical approach is also likely to have utility in “PheWAS” analysis - the identification of sets of traits associated with specific genetic variants, in that PheWAS can be thought of as the tracing of edges from a single starting point through the GRN, and this direction might also be explored if preferred.

This is an ambitious project; construction of GRNs has been a challenge for some time. However, recently established datasets providing genetic and expression data from multiple cell types from the same individuals make a solution newly feasible.

Informal enquiries to Chris Wallace are welcome: cew54@cam.ac.uk

More about our group: [http://chr1swallace.github.io](http://chr1swallace.github.io)

**Start date:** Easter Term (April) or Michaelmas Term (October) 2019

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](http://www.graduate.study.cam.ac.uk/applicant-portal) selecting course details MDBI22 PhD in Biostatistics

**Deadline for applications:** 3rd January 2019