Identifying the genetic determinants of translation rate in blood cells and characterising their relevance to human traits.

Supervisor: Ernest Turro (Department of Haematology), William Astle (MRC Biostatistics Unit) and Mattia Frontini (Department of Haematology)

The central dogma of molecular biology states that genetically encoded information flows from DNA to mRNA to protein. The flow from DNA to mRNA occurs through a process called transcription and the flow from mRNA to protein through a process called translation. Interestingly, measurements of transcript (i.e. mRNA) and protein abundance do not exhibit particularly strong correlation across genes. This lack of correlation is no doubt partly due to measurement error and partly due to heterogeneity in biological processes such as transcript and protein degradation across genes (1). Nevertheless, there remains the intriguing possibility that the imperfect correlation can be explained in part by heterogeneity in the translation rate, defined as the rate at which protein is generated from a unit quantity of transcripts with a particular RNA sequence, across genes. In addition to heterogeneity in the translation rate between genes within individuals, imperfect correlation between transcript and protein abundance across individuals within genes suggests that the translation rate for a given gene may vary across individuals. It is likely that this variation is partly due to genetic factors inherited in the DNA sequence.

Recent research has identified several mechanisms by which variation in the mRNA sequence can influence the translation rate, including variation that leaves the amino acid sequence of the corresponding protein product intact (2,3). Refining these mechanisms may improve understanding of disease aetiology (4) and aid the development of new therapies (5). The group of Dr Mattia Frontini in the Department of Haematology has generated datasets of transcript abundances measured by high-throughput RNA sequencing, and protein abundance levels measured by mass spectrometry, from four different blood cell types—platelets, monocytes, neutrophils and CD4+ T cells—in approximately 50 individuals (400 datasets in total). This distinctive dataset provides a unique opportunity to develop statistical methods for elucidating the genetic determinants of translation rates in blood cells and to discern the relevance of translation rates to a wide array of human quantitative and disease traits measured in 500,000 genotyped individuals in UK Biobank (6).

The student will develop methods to estimate the translation rate at the gene, isoform and individual level using paired RNA-seq and mass-spectrometric measurements of transcript and protein abundances. Genetic determinants of the translation rates will be identified using genetic association analysis. A general prediction model for translation rates will be developed. Using this model, rates will be imputed into 500,000 individuals from UK Biobank and tested for association with a wide array of measured traits.

This project will build on the experience of Drs William Astle and Ernest Turro in the domain of statistical genomics, notably using Bayesian modelling strategies (7,8,9) and the experience of Dr Mattia Frontini in blood genomics and cell biology (10,11). The successful candidate will have access to extensive computing facilities at the University's high performance computing cluster. Any potential findings will be amenable to rapid experimental follow-up in the laboratory.

References


**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 selecting course details MDB122 PhD in Biostatistics

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