Title: Integrated Bayesian approach to incorporate prior biological knowledge in the investigation of the effects of genetic variants on continuous disease traits

Miguel Pereira[1], John Thompson[2] and Cosetta Minelli[1]

[1]National Heart and Lung Institute, Imperial College London, UK
[2]Department of Health Sciences, University of Leicester, UK

Abstract

Genome-wide association studies (GWAS) have identified several single-nucleotide polymorphisms (SNPs) associated with complex diseases and traits. Usually, SNPs effects are estimated individually and their p-values adjusted for multiple testing since thousands or millions variants are studied. However, the estimated effect sizes for individual variants are typically small and regularization methods (e.g. LASSO) have been proposed to model SNPs jointly and improve statistical power by better exploiting the structure of the data. The Bayesian counterparts of classical regularization methods have the advantage of providing higher modelling flexibility and allowing the estimation of standard errors of the genetic effects. Moreover, Bayesian methods allow inclusion of prior biological knowledge in the analysis to further improve statistical power to detect genetic effects. The aim of our work is to integrate prior genetic knowledge in a Bayesian framework to improve the discovery of new variants in genome-wide association studies.

Currently, we are applying the Bayesian hierarchical model proposed by Yi *et al*. (*Plos Genet*. 2011) to jointly estimate the effects of SNPs grouped by Linkage Disequilibrium (LD) blocks. We have developed a systematic approach to retrieve prior biological knowledge from different databases, and we are now incorporating this information in the Bayesian model.