# Marginal SNP Effects

Matthew Sperrin

Department of Mathematics and Statistics, Lancaster University, UK

In a network of dense single nucleotide polymorphisms (SNPs) in the genome it is well known that there will be significant levels of linkage disequilibrium (LD) or correlation between SNPs that are in close proximity. LD patterns are complex and can even exist between SNPs over a long range. In a study on the effect of SNPs on some dichotomous phenotype, traditionally we would test each SNP to see whether there is evidence to suggest it is linked to the phenotype. Unfortunately, the linkage disequilibrium causes problems in this approach as the tests are correlated and consequently standard multiple testing procedures do not apply. The null distribution of the test statistics is unknown and is in general estimated via permutation testing. Another problem is that once a SNP is identified as potentially linked to the phenotype, we do not know whether this is caused by the direct causality of this SNP or that the causality is in fact a consequence of it being in LD with other causal SNPs. Hence, increasing the density of the SNP network seems to introduce more problems than it solves, under current methods.

Here we present a method in which we identify the *marginal effect* of each SNP — defined as the effect on the phenotype not explained via its LD with other SNPs. We therefore remove the problem caused by LD between the SNPs, and ensure that the effects we find are direct. We estimate, under some weak assumptions, the null distribution of each SNP as a function of the amount of information about this SNP that is available, which is related to the amount of LD it has with other SNPs.