

association in their available dataset conditional on the most associated SNP in the other [Nica et al., 2010]. When full data for both traits are available, colocalisation may be tested by examining whether coefficients from regressions of each trait against two or more SNPs are proportional, as they should be if those SNPs jointly tag a common causal variant [Wallace et al., 2012; Plagnol et al., 2009].

We show here that naïve application of both conditional and proportional colocalisation tests may result in substantially inflated type 1 errors, and explore reasons for that inflation. The inflation cannot be easily resolved for conditional tests, but we demonstrate two alternative approaches for proportional testing which result in unbiased inference. Finally, we apply these methods to colocalisation testing of 13 regions shown in Supplementary Table I which have been associated with one or both of the autoimmune thyroid diseases, Graves’ disease (GD) and Hashimoto’s thyroiditis (HT), using previously published dense genotyping data [Cooper et al., 2012].

## Methods

### Approaches to colocalisation testing

We begin by introducing some notation and setting out the details of the existing approaches to colocalisation testing that are explored in this paper. Assume two traits,  $Y$  and  $Y'$ , have been measured in distinct samples and evidence exists for association of both traits to some genetic region. Let the region be covered by  $p$  SNPs genotyped in both samples, with the genotype matrices denoted by  $X = (X_1, \dots, X_p)$  and  $X' = (X'_1, \dots, X'_p)$  respectively. Conditional approaches begin with identifying the most strongly associated SNPs for  $Y$  and  $Y'$ , SNPs  $k$  and  $k'$ , say, then examine whether there is any evidence for association between  $Y$  and SNP  $k$  conditional on SNP  $k'$ . The null hypothesis is therefore

$$H_0^{\text{cond}} : Y \perp\!\!\!\perp X_k | X_{k'}. \quad (1)$$

Concerned that LD would make interpretation of the conditional test difficult, Nica et al. [2010] extended the conditional method as follows. For every SNP  $j$  generate residuals  $R_j$  from a regression of  $Y$  against  $X_j$  and test the correlation of  $R_j$  and  $X_k$  using Spearman’s rank correlation