Covariate-modulated false discovery rates

Egil Ferkingstad, Arnoldo Frigessi, Gudmar Thorleifsson and Augustine Kong

Huge amounts of simultaneous comparisons are necessary to extract biological information from genetic and genomic data. Such tests are dependent, and the dependency structure is unknown, so that correcting for multiple testing is difficult. In an empirical Bayesian setting, the local false discovery rate (local FDR) is defined as the posterior probability that the null hypothesis is true given the data. We extend this methodology further and introduce the covariate-modulated false discovery rate (cmFDR), useful when an additional covariate is available, that a priori influences the probability for each null hypothesis being true. The cmFDR takes advantage of such covariate-modulated prior information to produce a new list of selected genes, which differs in length and order from the list produced by the local FDR. The cmFDR measures the posterior significance of each test conditionally on the covariate and the data, leading to greater power. We estimate the cmFDR with MCMC for an approximate model on p-values. The new method is applied to the analysis of expression quantitative trait loci (eQTL) data, where gene expression measurements from microarrays are combined with genetic linkage data. We also test for differentially expressed genes modulated by copy number alteration in breast cancer. Our method provides a simple way of integrating different data sets in a hypothesis testing context.