Adjusting for gene-specific covariates to improve false discovery rate estimation in RNA-seq analysis

This paper suggests a novel positive false discovery rate (pFDR) controlling method using a gene-specific covariate variable, such as gene length. We suppose the null probability depends on the covariate variable. In this context, we propose a rejection rule that accounts for heterogeneity among promising tests with low p-values, while accounting for different null probabilities. We establish a pFDR estimator for a given rejection rule by following Storey’s q-value framework. A condition on a type I error posterior probability is provided that equivalently characterizes our rejection rule. We also present a suitable procedure for selecting a tuning parameter through cross-validation that maximizes the expected number of hypotheses declared significant. A simulation study demonstrates that our method is comparable to or better than existing methods across a variety of realistic scenarios. In data analysis, we find support for our method’s premise that the null probability varies with a gene-specific covariate variable.