

Supplement to "Many phenotypes without many false discoveries: error controlling strategies for multi-trait association studies"

Christine B. Peterson, Marina Bogomolov, Yoav Benjamini, and Chiara Sabatti

Simulation with correlated test statistics

Here we provide a performance comparison of the various error control methods in a simulation setting where there is correlation among the test statistics within each hypothesis family. This is designed to reflect a scenario where there is dependence among phenotypes due to environmental or other factors, but the predictors are still independent. Specifically, we focus on a modified version of the simulation shown in Figure 3(A). We retain the dimensions used there, and so have $M=3000$ variants (hypothesis families) and $P=100$ phenotypes, with 60 variants each associated to 25 of the phenotypes. In the simulation corresponding to Figure 3(A), which utilizes independent test statistics, the null p -values are sampled from the uniform distribution on the $[0, 1]$ interval. This is equivalent to generating test statistics independently from the $\mathcal{N}(0, \sigma^2)$ distribution and calculating the two-tailed p -values under the $\mathcal{N}(0, \sigma^2)$ distribution. The non-null test statistics are generated independently from the $\mathcal{N}(2, \sigma^2)$ distribution, then the two-tailed p -values are calculated under the $\mathcal{N}(0, \sigma^2)$ distribution. To induce correlation of the test statistics across phenotypes, we modify the set-up as follows: the test statistics for each family are sampled from the multivariate normal distribution $\mathcal{N}_P(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where the mean vector $\boldsymbol{\mu}$ has entries set to 0 for null hypotheses and 2 for non-null hypotheses, where the position of the non-nulls is chosen at random, and the covariance matrix $\boldsymbol{\Sigma}$ is constructed as

$$\boldsymbol{\Sigma} = \sigma^2 \times \text{Toeplitz}(1, 0.5, 0.25, 0.1, 0, 0, \dots),$$

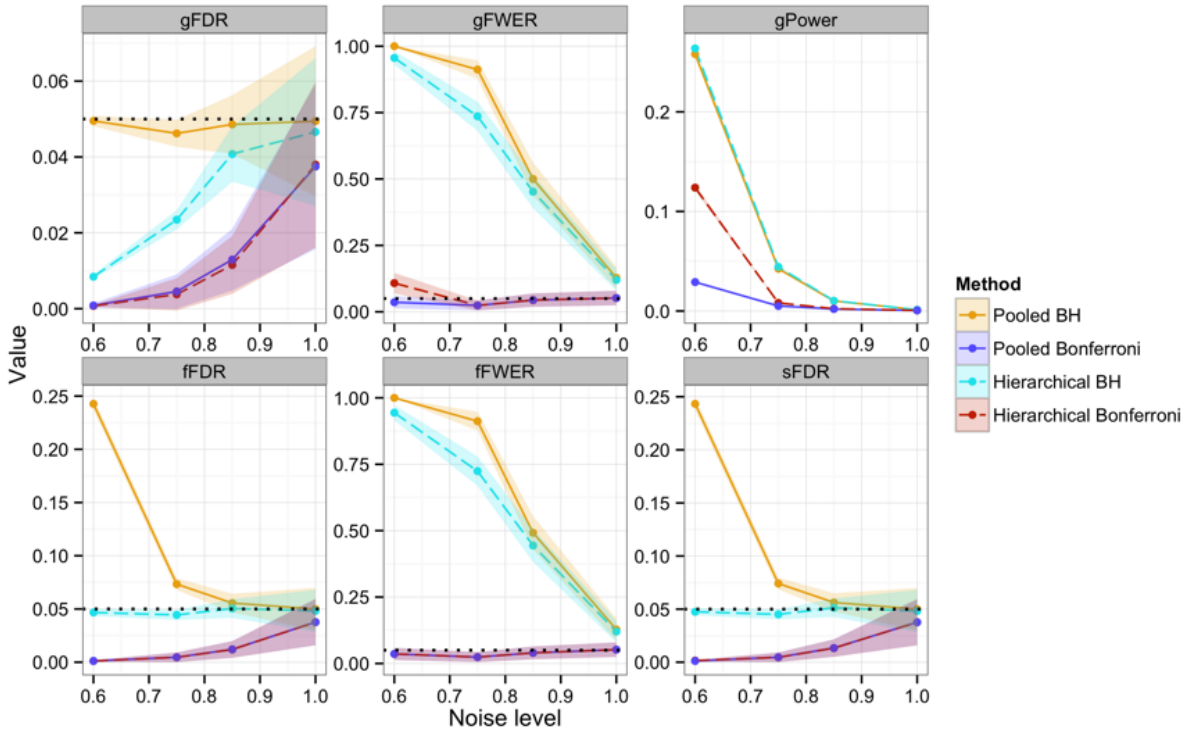


Figure S1: Error rates and power for four multiple-testing strategies. $M=3000$, $P=100$, 60 variants are associated with 25 traits each, and test statistics are correlated within hypothesis families. The solid lines show the average, the shaded areas represent the standard error over 250 iterations, and the dotted horizontal lines mark the level 0.05.

where `Toeplitz` represents the symmetric Toeplitz matrix of dimension $p \times p$ generated using the given vector. This choice of covariance matrix mimics the presence of spatial correlation across phenotypes. As shown in Figure S1, the results are very similar to those in Figure 3(A) in terms of the measured error rates and comparative performance of the different error controlling methods. This simulation demonstrates that the proposed hierarchical error control procedures do not break down in the presence of correlation among phenotypes.