servative. We come back to the issue of robustness of Procedure 3.2 in the Discussion Section 7.

GWAS simulation example. We simulated two GWAS from the simulator HAPGEN2 (Su et al., 2011). The two studies were generated from two samples of the HapMap project (The International HapMap Consortium, 2003), a sample of 165 Utah residents with Northern and Western European ancestry (CEU), and a sample of 109 Chinese in Metropolitan Denver, Colorado (CHD). In the CEU and CHD populations, respectively, 34 and 38 SNPs were set as disease SNPs with an increased multiplicative relative risk of 1.2, and 18 of the disease SNPs were common to both populations. Each study contained 4500 cases and 4500 referents. The linkage disequilibrium (LD) across SNPs, as measured for the samples in the HapMap project, was retained. Due to LD, the number of SNPs associated with the phenotype in each study was larger than the number of disease SNPs. In order to identify the SNPs in each study that are truly associated with the phenotype, the simulation of 4500 cases and 4500 controls from the population was repeated 11 times, and 11 p-values were produced per SNP. SNPs with Fisher's combined p-value (Loughin, 2004) below the Bonferroni threshold were considered to be truly associated with the disease. Our ground truth included 1355 and 1010 SNPs associated with the disease in the CEU and in the CHD population, respectively, out of which 274 SNPs were associated with the disease in both populations.

As a standard preprocessing step, we removed SNPs with minor allele frequency below 0.05, and thus the number of SNPs in the analysis was reduced from 1,387,466 to 887,362, on average, for the 11 pairs of studies. Our selection rule for Procedure 4.1 with parameters (w_1, q_1, q) was the BH procedure at level $w_1 q_1$ when the primary study was the CEU study, and at level $(1 - w_1)q_1$ when the primary study was the CHD study, since the potential set of SNPs to be discovered as having replicated