Supporting material to "Some Step-down Procedures Controlling the False Discovery Rate under Dependence"

Yongchao Ge and Stuart C. Sealfon and Terence P. Speed

Algorithm S1 presents the detailed computations for the step-down procedure of Section 4.2 of the main text. In Algorithm S1, the p-value p_i before Step 1 and the p-values $p_{i,b}$ in Step 2 are computed based on the the null distribution of T_i. This null distribution can be an appropriate Student t distribution when the data are normally distributed or can be computed by further resamplings.

Algorithm S1 Resampling algorithm for computing FDR adjusted p-values

---

Compute statistic t_i, and the p-value $p_i = P(|T_i| \geq |t_i| \mid H_i)$, $i = 1, \cdots, m$ for the observed data $\boldsymbol{x}$. Without loss of generality, label the p-values $p_1 \leq \cdots \leq p_m$.

For the b-th step, $b = 1, \ldots, B$, proceed as follows.

1. Compute the resampled data matrix $\boldsymbol{x}^b$, for example by randomly permuting the columns of matrix $\boldsymbol{x}$.

2. Compute the t-statistics $t_{1,b}, \ldots, t_{m,b}$ for each hypothesis on the resampled data $\boldsymbol{x}^b$ and then the p-values $p_{i,b} = P(|T_i| \geq |t_{i,b}| \mid H_i)$.

3. For each $i = 1, \ldots, m$, mimic the step-down procedure:

   (a) for $j = 1, \ldots, m - i + 1$, let $p_{(j)}^{i,b}$ be the j-th smallest member of $\{p_{i,b}, \ldots, p_{m,b}\}$;

   (b) define $R_{i,b}$ to be the unique integer k such that

   $$p_{(1)}^{i,b} \leq p_i, \ldots, p_{(k)}^{i,b} \leq p_{i+k-1} \text{ and } p_{(k+1)}^{i,b} > p_{i+k}.$$

4. Compute $f_{1,b} = I(R_{1,b} > 0)$ and, for $i = 2, \ldots, m$, compute $f_{i,b} = R_{i,b}/(R_{i,b} + i - 1)$.

Steps 1-4 are carried out B times and the adjusted p-values are estimated by $\check{p}_i = \sum_{b=1}^{B} f_{i,b}/B$, with monotonicity enforced by setting

$$\tilde{p}_1 \leftarrow \check{p}_1, \qquad \tilde{p}_i \leftarrow \max(\tilde{p}_{i-1}, \check{p}_i) \qquad \text{for } i = 2, \ldots, m.$$

---

m=1000 m1=100 rho=0 delta=1

m=1000 m1=500 rho=0 delta=1
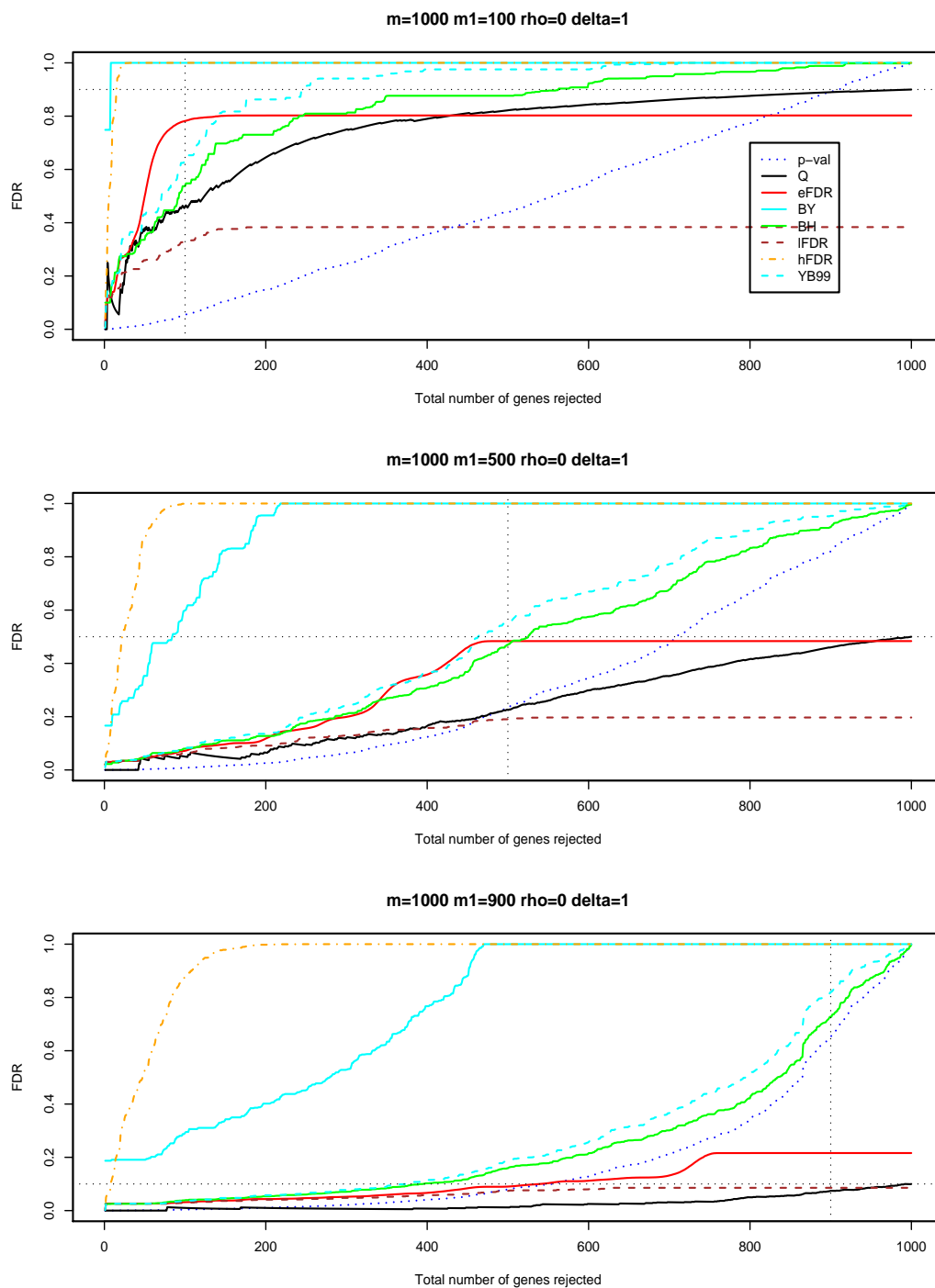

m=1000 m1=900 rho=0 delta=1

Figure S 1: Different FDR procedures. The independent case: $\rho = 0$ and $\delta = 1$: the dotted vertical line is $x = m_1$; the dotted horizontal line is $y = m_0/m$, the overall proportion of false null hypotheses. Different panels are for different values of m_1 (100, 500, 900).

m=1000 m1=100 rho=0.7 delta=1

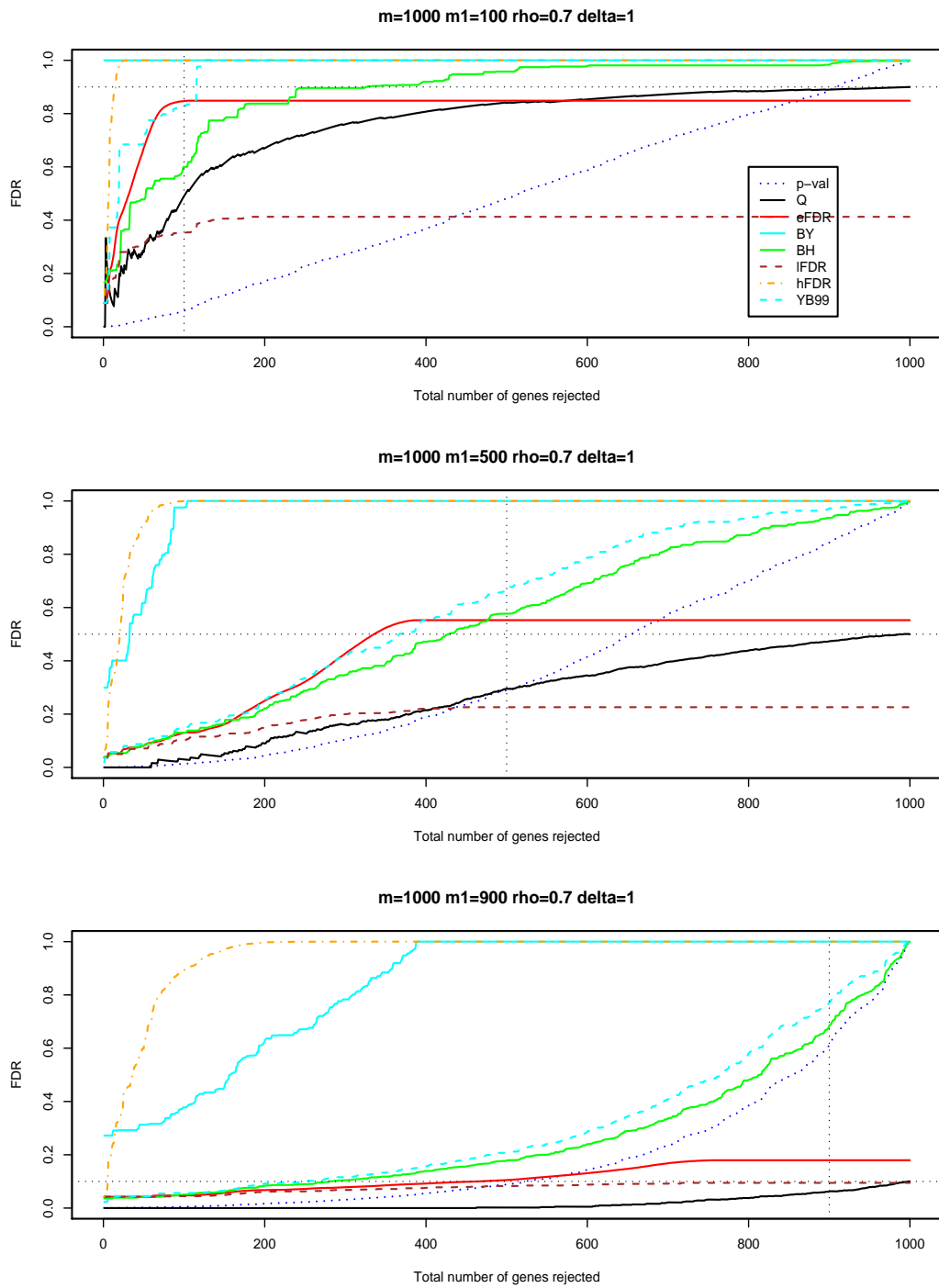**m=1000 m1=500 rho=0.7 delta=1**

m=1000 m1=900 rho=0.7 delta=1

Figure S2: The positively dependent case $\rho = 0.7$ and $\delta = 1$: different FDR procedures and different values of m_1 (100, 500, 900).
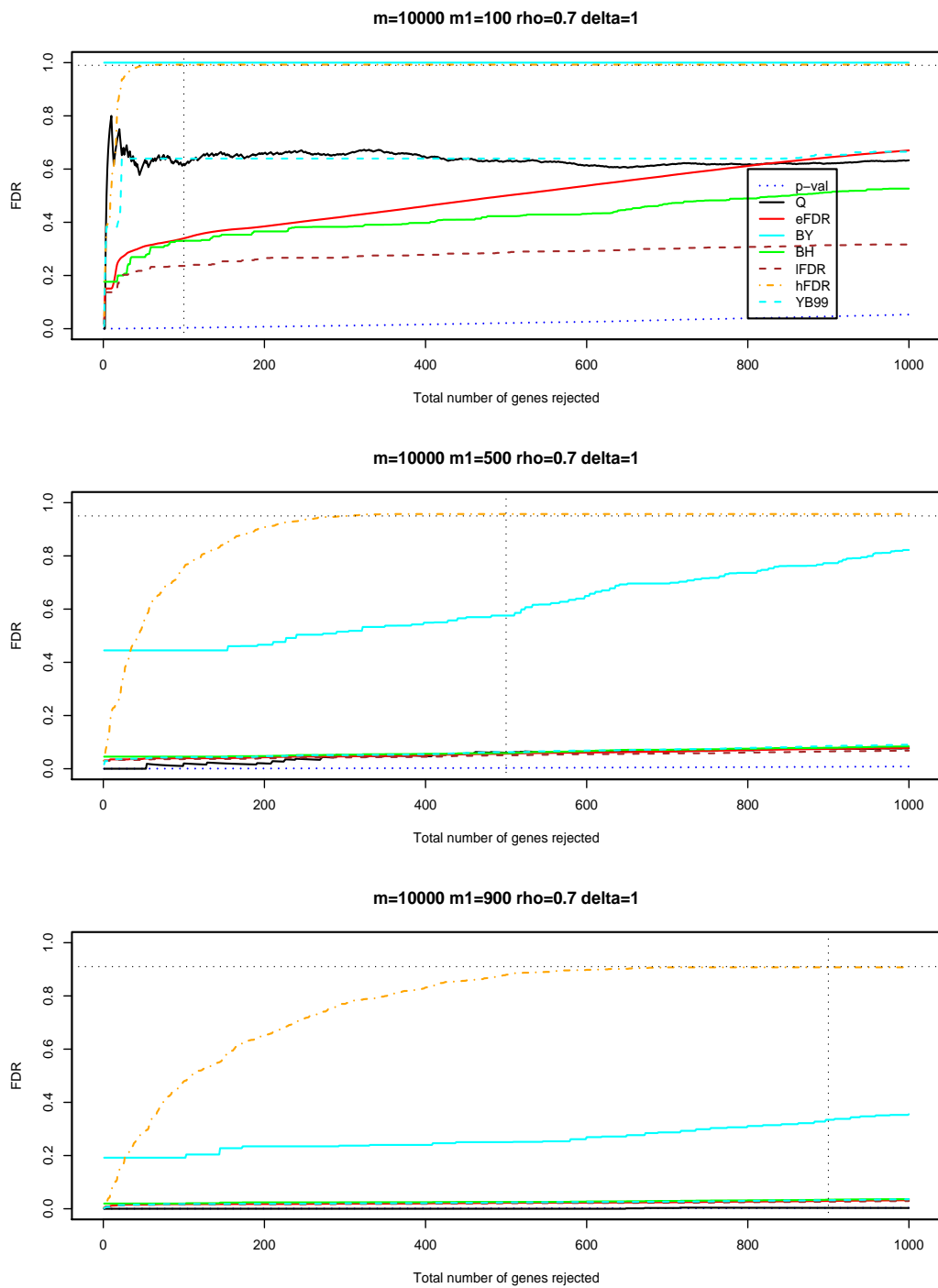
Figure S3: The same setup as in Figure S2 except that total number m of genes has increased from 1,000 to 10,000. We only plot the total number of genes rejected up to 1,000 among 10,000 genes.