SUPPLEMENTARY MATERIAL FOR REPLICABILITY ANALYSIS FOR GENOME-WIDE ASSOCIATION STUDIES

By Ruth Heller[†*] and Daniel Yekutieli[†]

Tel-Aviv University[†]

1. Proof of Proposition 3.1.

Proof. Since the result is for a single study i, for notational convenience we omit the subscript i in the following proof. Let $\psi_{OR}(z)$ be the indicator of whether $z \in \mathcal{Z}_{OR}$, and let $\psi(z)$ be the indicator of whether $z \in \mathcal{Z}$ for another rejection region that satisfies $Fdr(\mathcal{Z}) \leq q$. Straightforward calculus shows for every z

$$(1.1) \qquad \psi(z)(1 - fdr(z)/t(q)) \leq \psi_{OR}(z)(1 - fdr(z)/t(q)).$$

Taking expectations on both sides of equation (1.1),

$$\int [\psi(z)(1 - fdr(z)/t(q))]f(z)dz \leq \int [\psi_{OR}(z)(1 - fdr(z)/t(q))]f(z)dz,$$

we receive the following expression:

$$(1.2) \qquad P(\mathcal{Z})(1 - Fdr(\mathcal{Z})/t(q)) \leq P(\mathcal{Z}_{OR})(1 - Fdr(\mathcal{Z}_{OR})/t(q)).$$

Note that $q = E_f(fdr(z)|z \in \mathcal{Z}_{OR}) < E_f(t(q)|z \in \mathcal{Z}_{OR}) = t(q)$. Since $Fdr(\mathcal{Z}_{OR})$ is the expectation of $fdr(z)$ for $fdr(z) \leq t(q)$, and $q < t(q)$, it follows that $Fdr(\mathcal{Z}_{OR}) < t(q)$. Moreover, since $Fdr(\mathcal{Z}) \leq Fdr(\mathcal{Z}_{OR})$, it follows that $(1 - Fdr(\mathcal{Z})/t(q)) \geq (1 - Fdr(\mathcal{Z}_{OR})/t(q)) > 0$. Therefore, the right hand side of expression (1.2) is smaller than $P(\mathcal{Z}_{OR})(1 - Fdr(\mathcal{Z})/t(q))$ and item 1 follows.

In order to prove item 2, let $tdr(z) = 1 - fdr(z)$ be the true discovery rate. Straightforward calculus shows for every z

$$(1.3) \quad [1 - \psi(z)][1 - tdr(z)/(1 - t(q))] \leq [1 - \psi_{OR}(z)][1 - tdr(z)/(1 - t(q))]$$

Taking expectations on both sides of equation (1.3),

$$\int [1 - \psi(z)][1 - tdr(z)/(1 - t(q))]f(z)dz \leq \int [1 - \psi_{OR}(z)][1 - tdr(z)/(1 - t(q))]f(z)dz,$$

we receive the following expression:
(1.4)
$$[1 - P(\mathcal{Z})][1 - Fnr(\mathcal{Z})/(1 - t(q))] \leq [1 - P(\mathcal{Z}_{OR})][1 - Fnr(\mathcal{Z}_{OR})/(1 - t(q))]$$

Since $fdr(z) > t(q)$ for $z \notin \mathcal{Z}_{OR}$, it follows that $1 - tdr(z)/(1 - t(q)) > 0$ for $z \notin \mathcal{Z}_{OR}$, and therefore that $1 - Fnr(\mathcal{Z}_{OR})/(1 - t(q)) > 0$. Combining this observations with the fact from item 1 that $1 - P(\mathcal{Z}) \geq 1 - P(\mathcal{Z}_{OR})$, the RHS of equation (1.4) can be bounded above by $[1 - P(\mathcal{Z})][1 - Fnr(\mathcal{Z}_{OR})/(1 - t(q))]$. It thus follows that $1 - Fnr(\mathcal{Z})/(1 - t(q)) \leq 1 - Fnr(\mathcal{Z}_{OR})/(1 - t(q))$, proving item 2. □

2. Testing normal means. In this section we give simple examples that demonstrate that the rejection region for replicability analysis is very different than for an analysis to discover associations, and also that the optimal rejection regions may be far larger than a rejection region based on p-values. In this section only, for simplicity, we assume that each hypothesis has only two states: the null state with zero expectation, and the non-null state with positive expectation.

2.1. *Comparison of Bayes FDR for optimal and p-value based Bayesian analysis.*

EXAMPLE 2.1. *For $n = 2$ studies, suppose the marginal z-score density in the two studies is $N(0, 1)$ under the no-association null hypothesis, and under the alternative positive association hypotheses the z-score density is $N(\mu_1, 1)$ in the first study and $N(\mu_2, 1)$ in the second study. Thus the joint z-score density is*

$$\begin{aligned} f(z_1, z_2) &= \pi(0, 0)\phi(z_1)\phi(z_2) + \pi(0, 1)\phi(z_1)\phi(z_2 - \mu_2) \\ &\quad + \pi(1, 0)\phi(z_1 - \mu_1)\phi(z_2) + \pi(1, 1)\phi(z_1 - \mu_1)\phi(z_2 - \mu_2), \end{aligned}$$

where $\phi(z)$ is the standard Normal density. For $\vec{h} \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, the conditional probability that $\vec{H} = (h_1, h_2)$ given (z_1, z_2) is

$$\Pr(\vec{H} = \vec{h}|\ (z_1, z_1)) = \frac{\pi(\vec{h})\phi(z_1 - I(h_1 = 1)\mu_1)\phi(z_2 - I(h_2 = 1)\mu_2)}{f(z_1, z_2)}.$$

The local Bayes fdr for testing H_{NA}^0 and H_{NR}^0, respectively, is

$$fdr_{NA}(z_1, z_2) = \Pr(\vec{H} = (0, 0)|\ (z_1, z_1))$$

and

$$fdr_{NR}(z_1, z_2) = \Pr(\vec{H} = (0, 0)|\ (z_1, z_2)) + \Pr(\vec{H} = (0, 1)|\ (z_1, z_2)) + \Pr(\vec{H} = (1, 0)|\ (z_1, z_2)).$$

TABLE 1

The probability of the optimal and of the p-value based rejection regions, for various Bayes FDR levels q and two configurations of $\pi = (\pi(0,0), \pi(0,1), \pi(1,0), \pi(1,1))$.

Null	$\pi(0,0), \pi(0,1), \pi(1,0), \pi(1,1)$	Rejection region	$q = 0.05$	$q = 0.20$
H_{NR}^0	$(0.80, 0.08, 0.08, 0.04)$	$\mathcal{Z}_{OR}$	0.0234	0.0417
		p-value: Bayes	0.0230	0.0410
		BHY09	0.0028	0.0145
H_{NA}^0	$(0.80, 0.08, 0.08, 0.04)$	$\mathcal{Z}_{OR}$	0.1498	0.2230
		p-value: Bayes	0.1417	0.2182
		BHY09	0.1334	0.2007
H_{NA}^0	$(0.88, 0.12, 0.00, 0.00)$	$\mathcal{Z}_{OR}$	0.0855	0.1355
		p-value: Bayes	0.0621	0.1178
		BHY09	0.0563	0.1050

We compared the optimal rejection region and the rejection region based on p-values for the Bayesian analysis. The p-values for H_{NA}^0 and H_{NR}^0 were, respectively, the p-values of the Fisher combined (right-sided) p-values, and the maximum of the two studies p-values. Specifically, for $\vec{Z} = (Z_1, Z_2)$, let $P_1 = 1 - \Phi(Z_1)$ and $P_2 = 1 - \Phi(Z_2)$. The p-value for testing no-association was $P^{NA} = 1 - F_{\chi_4^2}(-2(log(P_1) + log(P_2)))$, and the p-value for testing no-replication was $P^{NR} = \max(P_1, P_2)$.

For $\mu_1 = \mu_2 = 3$, let $\pi(0,0) = 0.80, \pi(0,1) = \pi(1,0) = 0.08$, and $\pi(1,1) = 0.04$. Figure 1 shows the rejection region boundaries for the optimal rejection region (solid) and the p-values based rejection region (dashed) for testing H_{NA}^0 (top left) and H_{NR}^0 (top right). Clearly, the rejection regions are much larger for detecting associations than for detecting replicability.

For $\pi(0,0) = 0.88, \pi(0,1) = 0.12$, Figure 1 (bottom) shows the rejection regions when testing H_{NA}^0. The hypothesis H_{NR}^0 is not tested, since the local Bayes FDR of no replicability is one, and there does not exist a region with Bayes FDR at most $q < 1$. The difference between the optimal rejection region and the rejection region based on p-values is much larger in this configuration than in the previous configuration. Specifically, the optimal rejection region is only determined by the z-score of the second study, Z_2.

Table 1 shows the probability of the rejection regions for the no association and no replicability null hypotheses. The probabilities of the rejection regions to discover replicability are much smaller than for discovering associations. Moreover, the probabilities of the optimal rejection regions are larger than for the p-value based region, and the differences between the probabilities of the regions are larger for configuration $\pi(0,0) = 0.88, \pi(0,1) = 0.12$ than for $\pi(0,0) = 0.80, \pi(0,1) = \pi(1,0) = 0.08$, and $\pi(1,1) = 0.04$. See Section 2.2 for a discussion of the results for the method of BHY09.
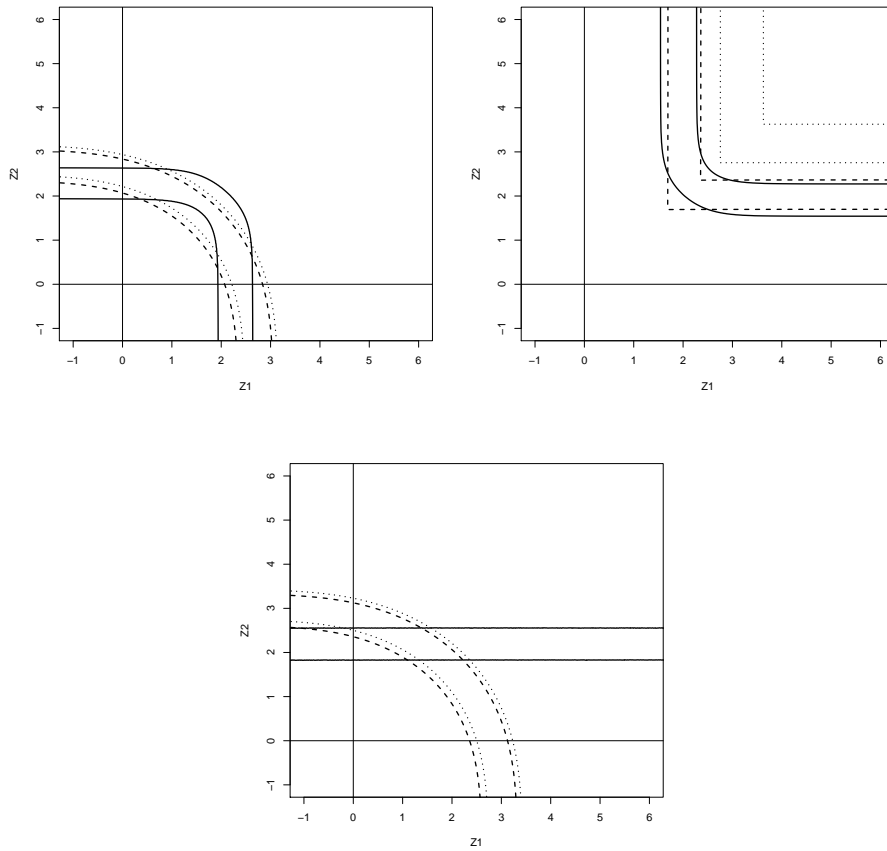
FIG 1. *Optimal (solid curves), p-value based (dashed curves) rejection regions boundaries for bayes FDR levels $q \in \{0.20, 0.05\}$, as well as the rejection region for the analysis of BHY09 (dotted curves) for FDR levels $q \in \{0.20, 0.05\}$, in configuration $\pi((0,0)) = 0.80, \pi((0,1)) = \pi((1,0)) = 0.08$, and $\pi((1,1)) = 0.04$ of the test of H_{NA}^0 (top left) and H_{NR}^0 (top right) , and in configuration $\pi((0,0)) = 0.88, \pi((0,1)) = 0.12$ for the test of H_{NA}^0 (bottom). The further the boundary is from (0,0) the smaller the value of q.*

The following example illustrates the large loss of power due to a non-optimal choice of rejection region that can occur when more than two studies are available.

EXAMPLE 2.2. *For $n = 6$ studies, let $\pi((0,0,0,0,0,0)) = 0.90$ and $\pi((0,0,0,0,0,1)) = 0.10$. Thus the first five z-scores $Z_1 \cdots Z_5$ are $N(0,1)$. The sixth z-score Z_6 is $N(0,1)$ with probability 0.9 and $N(3,1)$ with probability 0.1. Similar to the setting $(\mu_1, \mu_2) = (0,3)$ in Example 2.1, the p-value based rejection region for testing H_{NA}^0 is very different than the optimal rejection region, which is only based on Z_6. For a Bayes FDR of $q = 0.05$, the probability of the optimal rejection region was 0.066, and the probability of the p-value based rejection region was 0.012.*

2.2. *Comparison of Bayes FDR for p-value based Bayesian analysis and for the BH procedure.* In Example 2.1, the dotted curve in Figure 1 shows the rejection region using the BH procedure, as suggested in BHY09. While the rejection region is only slightly smaller than that of the p-value based Bayesian rejection region for testing for no association (top left and bottom), it is much smaller for testing for no replicability (top right). We shall explain why these differences arise.

In the two group model, when the rejection region is based on the tails of the z-scores $\mathcal{Z} = \{z : z \leq t(q)\}$ which are equivalent to one-sided p-values, there is a strong connection between empirical Bayes estimation of the Bayes FDR and the frequentist BH procedure for FDR control, as noted by Efron and Tibshirani (2002) and Storey (2002). If the jth p-value in study i is $p_{ij} = \Phi(z_{ij})$, then the BH rule rejects all hypotheses with z-scores that satisfy the following inequality:

$$(2.1) \qquad \widehat{Fdr}_{i,BH}(z_{i(j)}) = \max_{l \geq j} \Phi(z_{i(l)})/(l/M) \leq q,$$

where $z_{i(j)}$ is the jth largest z-score in study i. Since j/M is the empirical distribution of $\mathcal{Z}$ for the rejection region $\mathcal{Z}_j = \{z : z \leq z_{i(j)}\}$, then if we set $\pi_0(i)$ conservatively to be one, the BH procedure coincides with the procedure that chooses the largest $\mathcal{Z}_j$ so that the estimated $Fdr_i(\mathcal{Z}_j)$ is at most q. Specifics follow. The rejection region of the BH procedure is $\mathcal{Z}_{\hat{p}_{i,BH}} = \{\vec{z} : P_{ij} \leq \hat{p}_{BH}\}$, where $\hat{p}_{i,BH} = \sup\{p : \widehat{Fdr}_{i,BH}(p) \leq q\}$ and

$$(2.2) \qquad \widehat{Fdr}_{i,BH}(p) = \frac{p}{|\{j : p_{ij} \leq p\}|/M}.$$

The Bayes FDR of $\mathcal{Z}_p = \{z_{ij} : P_{ij} \leq p\}$ is

$$(2.3) \qquad Fdr_i(\mathcal{Z}_p) = \Pr(H_{ij} = 0 | P_{ij} \in \mathcal{Z}_p) = \frac{\pi_0(i) \Pr(P_{ij} \leq p | H_{ij} = 0)}{\Pr(P_{ij} \in \mathcal{Z}_p)}$$

Comparing (2.2) with (2.3), as the denominator of (2.2) is the empirical distribution of the event in the denominator of (2.3), if P_{ij} is $U[0,1]$ under the null hypothesis, the Fdr estimator in (2.2) is too large by a factor of $1/\pi_0(i)$. If P_{ij} is stochastically greater than $U[0,1]$, the Fdr estimator in (2.2) may be greatly over-conservative.

Similarly, for the null hypothesis $\mathcal{H}^0_{NA}$, the conservative factor is $1/\pi(\vec{0})$, since the rejection region of the BH procedure is $\mathcal{Z}_{\hat{p}^{NA}_{BH}} = \{\vec{z}_j : p^{NA}_j \le \hat{p}^{NA}_{BH}\}$, where $\hat{p}^{NA}_{BH} = \sup\{p : \widehat{Fdr}^{NA}_{BH}(p) \le q\}$ and

$$(2.4) \qquad \widehat{Fdr}^{NA}_{BH}(p) = \frac{p}{|\{\vec{z}_j : p^{NA}_j \le p\}|/M},$$

and the Bayes FDR of $\mathcal{Z}_p = \{\vec{z} : p^{NA}_j \le p\}$ is

$$(2.5) \qquad Fdr^{NA}(\mathcal{Z}_p) = \Pr(\mathcal{H}^0_{NA}|\ \vec{P}_j \in \mathcal{Z}_p) = \frac{\pi(\vec{0})\Pr(P^{NA} \le p|\vec{H} = \vec{0})}{\Pr(\vec{z}_j \in \mathcal{Z}_p)}$$

However, $\mathcal{H}^0_{NR}$ is a composite null hypothesis and therefore the conservativeness of the BH procedure is far greater. The rejection region of the BH procedure is $\mathcal{Z}_{\hat{p}^{NR}_{BH}} = \{\vec{z}_j : p^{NR}_j \le \hat{p}^{NR}_{BH}\}$, where $\hat{p}^{NR}_{BH} = \sup\{p : \widehat{Fdr}^{NR}_{BH}(p) \le q\}$ and

$$(2.6) \qquad \widehat{Fdr}^{NR}_{BH}(p) = \frac{p}{|\{\vec{z}_j : p^{NR}_j \le p\}|/M}.$$

The Bayes FDR of $\mathcal{Z}_p = \{\vec{z} : p^{NR}_j \le p\}$ is
(2.7)
$$Fdr^{NR}(\mathcal{Z}_p) = \Pr(\mathcal{H}^0_{NR}|\ \vec{z}_j \in \mathcal{Z}_p) = \frac{\sum_{\vec{h} \in \mathcal{H}^0_{NR}} \pi(\vec{h})\Pr(P^{NR} \le p|\vec{H} = \vec{h})}{\Pr(\vec{z}_j \in \mathcal{Z}_p)}$$

Comparing (2.6) with (2.7), as the denominator of (2.6) is the empirical distribution of the event in the denominator of (2.7), the conservatism of the BH procedure follows from the differences in the numerators of these two expressions. The BH procedure is conservative since P^{NR}_j is stochastically greater than $U[0,1]$ for $\vec{H} \in \mathcal{H}_0$, especially for $\vec{H} = \vec{0}$. Therefore, the numerator in (2.6) is much larger than the numerator in (2.7) when $\pi(\vec{0})$ is large.

For Example 2.1, Table 1 shows the probability of the BH rejection regions for $(\mu_1, \mu_2) = (3,3)$. Figure 2 shows $\widehat{Fdr}^{NA}_{BH}(p)$ versus $Fdr^{NA}(\mathcal{Z}_p)$ and $\widehat{Fdr}^{NR}_{BH}(p)$ versus $Fdr^{NR}(\mathcal{Z}_p)$. For testing H^0_{NA}, $\widehat{Fdr}_{BH}$ was overly conservative by a factor of 1.25. Therefore the rejection region with $\widehat{Fdr}_{BH} = 0.05$
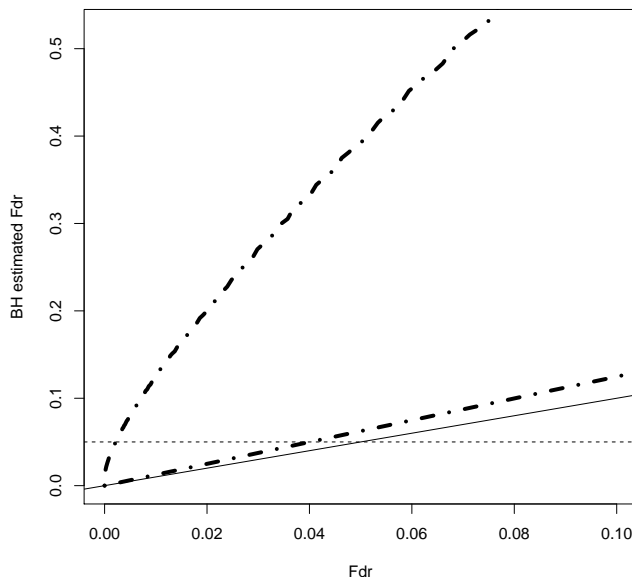
FIG 2. *The Bayes FDR value q (solid line), $\widehat{Fdr}_{BH}$ for the test of H_{NA}^0 (bottom dash-dot line) and for the test of H_{NR}^0 (top dash-dot line). The horizontal dashed line is at level 0.05, and it intersects the bottom and top dash-dot lines at $(Fdr, \widehat{Fdr}_{BH})$ values (0.05,0.05) and (0.0022, 0.05), respectively.*

actually had $Fdr = 0.04$, and a rejection probability of 0.133, while the rejection probability was 0.1417 for the p-value based rejection region with $Fdr = 0.05$. For testing H_{NR}^0, the rejection region with $\widehat{Fdr}_{BH} = 0.05$ actually had $Fdr = 0.0022$, and the rejection probability was only 0.0028. For comparison, the rejection probability was 0.0230 for the p-value based rejection region with $Fdr = 0.05$.

3. Computation of $f(\vec{z})$. The locfdr package estimates $f(z_{ij}|\ H_{ij} = 0)$ and $\Pr(H_{ij} = 0)$ in addition to $f(z_{ij})$, and then derives $f(z_{ij}|\ H_{ij} \neq 0)$ through the relation

$$f(z_{ij}) = f(z_{ij}|\ H_{ij} = 0) \cdot \Pr(H_{ij} = 0) + f(z_{ij}|\ H_{ij} \neq 0) \cdot \{1 - \Pr(H_{ij} = 0)\}.$$

In replicability analysis, that considers

$$\Pr(\vec{H}_j = \vec{h}|\ \vec{z}_j) = \frac{f(\vec{z}_j,\ \vec{H}_j = \vec{h})}{f(\vec{z}_j)}$$

(3.1)
$$= \frac{\{\Pi_{i=1}^n f(z_{ij}|\ H_{ij} = h_i)\} \cdot \Pr(\vec{H}_j = \vec{h})}{f(\vec{z}_j)}$$

for $\vec{h} \neq \vec{0}$, it is also necessary to specify $f(z_{ij}|\ H_{ij} = -1)$ and $f(z_{ij}|\ H_{ij} = 1)$.

If H_{ij} are independent then the components of $\vec{z}_j$ are also independent and thus the locfdr estimates of the marginal z-score densities are sufficient for computing

$$\Pr(\vec{H}_j = \vec{0}|\ \vec{z}_j) = \frac{\Pi_{i=1}^n \{f(z_{ij}|\ H_{ij} = 0) \cdot \Pr(H_{ij} = 0)\}}{\Pi_{i=1}^n f(z_{ij})}.$$

However, if the components of $\vec{H}_j$ are dependent then specifying $f(z_{ij}|\ H_{ij} = -1)$ and $f(z_{ij}|\ H_{ij} = 1)$ is necessary for computing $\Pr(\vec{H}_j = \vec{0}|\ \vec{z}_j)$, as illustrated in the example below.

EXAMPLE 3.1. *Assume that either* $\vec{H}_j = \vec{0}$ *or* $\vec{H}_j \in \mathcal{H}^1$, *for* $\mathcal{H}^1 = \{-1, 1\}^n$, *and let* $\Pr(\vec{H}_j = \vec{0}) = \pi_0$. *Therefore,* $f(\vec{z}_j, \vec{H}_j = \vec{0}) = \{\Pi_{i=1}^n f(z_{ij}|\ H_{ij} = 0)\} \cdot \pi_0$. *Since* $f(\vec{z}_j) = f(\vec{z}_j, \vec{H}_j = \vec{0}) + f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1)$, *to compute* $\Pr(\vec{H}_j = \vec{0}|\ \vec{z}_j)$ *we need* $f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1)$. *In general,*

$$f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1) = \sum_{\vec{h} \in \mathcal{H}^1} f(\vec{z}_j, \vec{H}_j = \vec{h})$$

$$= \sum_{\vec{h} \in \mathcal{H}^1} f(\vec{z}_j|\ \vec{H}_j = \vec{h}) \cdot \Pr(\vec{H}_j = \vec{h})$$

(3.2)
$$= \sum_{\vec{h} \in \mathcal{H}^1} \{\Pi_{i=1}^n f(z_{ij}|\ H_{ij} = h_i)\} \cdot \Pr(\vec{H}_j = \vec{h}).$$

If the components of $\vec{H}_j$ *were independent conditional on* $\vec{H}_j \in \mathcal{H}^1$, *then*

(3.3)
$$\begin{aligned} f(\vec{z}_j, \vec{H}_j \in \mathcal{H}^1) &= f(\vec{z}_j|\ \vec{H}_j \in \mathcal{H}^1) \cdot \Pr(\vec{H}_j \in \mathcal{H}^1) \\ &= \{\Pi_{i=1}^n f(z_{ij}|\ H_{ij} \neq 0)\} \cdot (1 - \pi_0). \end{aligned}$$

Note that to compute (3.2) it is necessary to estimate $f(z_{ij}|\ H_{ij} = -1)$ *and* $f(z_{ij}|\ H_{ij} = 1)$ *and that if* H_{ij} *are independent conditional on* $\vec{H}_j \in \mathcal{H}^1$ *then expressions (3.2) and (3.3) are the same, but for large* n *and highly*

dependent H_{ij} they may be very different. To see this, we further assume $\Pr(\vec{H}_j = (1,\ldots,1)) = (1-\pi_0)/2$ *and* $\Pr(\vec{H}_j = (-1,\ldots,-1)) = (1-\pi_0)/2$, *and consider* $\vec{z} = (z_{1j}\cdots z_{nj})$ *with* $0 < z_{ij}$ *for which* $f(z_{ij}|\ H_{ij}=-1) <<$ $f(z_{ij}|\ H_{ij}=0)$ *and* $f(z_{ij}|\ H_{ij}=0) << f(z_{ij}|\ H_{ij}=1)$. *Since* $f(z_{ij}|\ H_{ij}=-1) << f(z_{ij}|\ H_{ij}=1)$, *expression (3.2) can be approximated as follows:*

$$\sum_{\vec{h}\in\mathcal{H}^1} \{\Pi_{i=1}^n f(z_{ij}|\ H_{ij}=h_i)\}\cdot\Pr(\vec{H}_j=\vec{h})$$

(3.4)
$$\approx \{\Pi_{i=1}^n f(z_{ij}|\ H_{ij}=1)\}\cdot(1-\pi_0)/2.$$

Furthermore since $\Pr(H_{ij}=1|\ H_{ij}\neq 0) = 1/2$ *and*

$$\begin{aligned}f(z_{ij}|\ H_{ij}\neq 0) &= f(z_{ij}|\ H_{ij}=-1)\cdot\Pr(H_{ij}=-1|\ H_{ij}\neq 0)\\ &+ f(z_{ij}|\ H_{ij}=1)\cdot\Pr(H_{ij}=1|\ H_{ij}\neq 0),\end{aligned}$$

then $f(z_{ij}|\ H_{ij}=1)/f(z_{ij}|\ H_{ij}\neq 0)\approx 2$. *Thus expression (3.4) is* $2^{(n-1)}$ *larger than expression (3.3). Since* $f(z_{ij}|\ H_{ij}=0) << f(z_{ij}|\ H_{ij}=1)$, *it follows that* $f(\vec{z}_j)\approx f(\vec{z}_j,\vec{H}_j\in\mathcal{H}^1)$. *As the denominator of* $\Pr(\vec{H}_j=\vec{0}|\ \vec{z}_j)$ *is approximately* $f(\vec{z}_j,\vec{H}_j\in\mathcal{H}^1)$ *then in this case* $\Pr(\vec{H}_j=\vec{0}|\ \vec{z}_j)$ *is* $2^{(n-1)}$ *smaller then it would have been if* H_{ij} *were independent conditional on* $\vec{H}_j\in\mathcal{H}^1$.

4. The EM algorithm. The observed data are z-scores $\vec{z}_1,\ldots,\vec{z}_M$ and the missing values are $\vec{H}_1,\ldots,\vec{H}_M$. The complete likelihood for SNP j is

$$L_c(\vec{\pi};\vec{z}_j,\vec{f},\vec{H}_j) = f(\vec{z}_j|\vec{H}_j)\pi(\vec{H}_j).$$

The composite complete likelihood for all the SNPs is

$$\Pi_{j=1}^M L_c(\vec{\pi};\vec{z}_j,\hat{f},\vec{H}_j) = \Pi_{j=1}^M f(\vec{z}_j|\vec{H}_j)\pi(\vec{H}_j).$$

E step. In the E step we calculate the expected value of the log composite likelihood function, with respect to the conditional distribution of H given $\vec{z}$ under the current estimate of the parameters, $\vec{\pi}^{(t)}$:

$$\begin{aligned}Q(\vec{\pi}|\ \vec{\pi}^{(t)}) &= E_{H|\vec{z},\vec{\pi}^{(t)}}[log\{\Pi_{j=1}^M f(\vec{z}_j|\vec{H}_j)\cdot\pi(\vec{H}_j)\}]\\ &= \sum_{j=1}^M\sum_{\vec{h}\in\mathcal{H}}\Pr(\vec{H}_j=\vec{h}|\vec{z}_j,\vec{\pi}^{(t)})[\log f(\vec{z}_j|\vec{H}_j=\vec{h}) + \log\{\pi(\vec{h})\}]\\ &= \sum_{j=1}^M\sum_{\vec{h}\in\mathcal{H}}\Pr(\vec{H}_j=\vec{h}|\vec{z}_j,\vec{\pi}^{(t)})\log f(\vec{z}_j|\vec{H}_j=\vec{h})\end{aligned}$$

(4.1)
$$+ \sum_{j=1}^M\sum_{\vec{h}\in\mathcal{H}}\Pr(\vec{H}_j=\vec{h}|\vec{z}_j,\vec{\pi}^{(t)})\log\{\pi(\vec{h})\}$$

where

$$\Pr(\vec{H}_j = \vec{h}|\vec{z}_j, \vec{\pi}^{(t)}) = \frac{f(\vec{z}_j|\vec{h})\pi^{(t)}(\vec{h})}{\sum_{\vec{h'} \in \mathcal{H}} f(\vec{z}_j|\vec{h'},)\pi^{(t)}(\vec{h'})}$$

M step. Find $\vec{\pi}^{(t+1)}$ that maximizes $Q(\vec{\pi}|\ \vec{\pi}^{(t)})$. Since the second sum in equation (4.1) has the same form as the log-likelihood for the multinomial distribution, it follows that

$$\pi^{(t+1)}(\vec{h}) = \frac{\sum_{j=1}^{M} \Pr(\vec{H}_j = \vec{h}|\vec{z}_j, \vec{\pi}^{(t)})}{\sum_{\vec{h'} \in \mathcal{H}} \sum_{j=1}^{M} \Pr(\vec{H}_j = \vec{h'}|\vec{z}_j, \vec{\pi}^{(t)})}.$$

The updated parameters are $\vec{\pi}^{(t+1)} = \{\pi^{(t+1)}(\vec{h}) : \vec{h} \in \mathcal{H}\}$.

starting value $\pi^{(0)}$. As starting values, we recommend using values constrained to satisfy $\hat{\pi}_0(i) = \sum_{\{\vec{h} \in \mathcal{H}\} \cap \{h_i=0\}} \pi^{(0)}(\vec{h})$. Such a starting position will provide a good initial estimate of the non-null densities in the E step. Specifically, given estimates $\hat{\pi}_0(i), i = 1, \ldots, n$, we suggest as starting values

$$\pi^{(0)}(\vec{h}) = \Pi_{i=1}^{n} \hat{\pi}_{h_i}(i),$$

where $\hat{\pi}_1(i) = \hat{\pi}_{-1}(i) = (1 - \hat{\pi}_0(i))/2$.

Updating of $f(\vec{z}_j|\vec{h})$. After the EM converged to a new estimate, the estimated fraction of null hypotheses in each study can be extracted: $\hat{\pi}_0^{(T)}(i) = \sum_{\{\vec{h} \in \mathcal{H}\} \cap \{h_i=0\}} \pi^{(T)}(\vec{h})$, where T is the number of steps till convergence of the EM. A modified estimate of $f_{i,1}$ and $f_{i,-1}$ can then be computed using the new estimates $\hat{\pi}_0^{(T)}(i), i = 1, \ldots, n$, if these estimates are different from the starting values $\hat{\pi}_0(i), i = 1, \ldots, n$. These modified estimates can now be used to recompute $f(\vec{z}_j|\vec{h})$. Next, the EM can be repeated with the new estimated conditional densities. This iterative process should end when the new estimates of $\hat{\pi}_0^{(T)}(i), i = 1, \ldots, n$, are almost the same as the starting values of the EM.

5. Replicability analysis of T2D GWA studies. Figure 3 shows the empirical z-scores, as well as the estimated conditional densities, for each of the six studies, as outputted from the locfdr package.

The table below gives the list of the 219 SNPs with replicated associations, as discovered by the empirical Bayes analysis, sorted by positions on the chromosome. The positions were found by NCBI build GRCh37.p5 reference assembly, and they were mapped to nearby genes by dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/dbSNP.cgi?list=rslist).

The table shows the estimated Bayes FDR for replicability analysis as well as for the analysis to discover association, and the adjusted p-values from the corresponding analysis of BHY09 based on all six available studies.

	chr	pos	gene	Empirical Bayes Fdr		BHY09 adjusted p-values	
				Repl.	Assoc.	Repl.	Assoc.
rs10923931	1	120517959	NOTCH2	1.34e-02	2.70e-03	1.00e+00	3.45e-04
rs6442307	3	12143355	SYN2	4.43e-02	2.74e-02	1.00e+00	4.89e-02
rs11715886	3	12147236	SYN2	4.40e-02	2.73e-02	1.00e+00	4.89e-02
rs4488811	3	12182028	SYN2	3.74e-02	2.36e-02	1.00e+00	4.21e-02
rs11721223	3	12185160	SYN2	3.67e-02	2.35e-02	1.00e+00	4.11e-02
rs11708978	3	12188495	SYN2	3.64e-02	2.34e-02	1.00e+00	4.06e-02
rs6792867	3	12189900	SYN2	3.77e-02	2.37e-02	1.00e+00	4.02e-02
rs7629805	3	12192394	SYN2	4.79e-02	3.15e-02	1.00e+00	5.10e-02
rs10433537	3	12198485	SYN2,TIMP4	3.60e-02	2.33e-02	1.00e+00	3.86e-02
rs13070993	3	12217797	SYN2	3.70e-02	2.35e-02	1.00e+00	3.69e-02
rs11720578	3	12267084	non-coding	4.33e-02	2.68e-02	1.00e+00	4.81e-02
rs13071168	3	12275447	non-coding	1.39e-02	1.24e-02	1.00e+00	1.53e-02
rs11709119	3	12276493	non-coding	4.14e-02	2.97e-02	1.00e+00	1.56e-02
rs17036101	3	12277845	non-coding	1.47e-02	1.26e-02	1.00e+00	1.56e-02
rs1562040	3	12285405	non-coding	1.43e-02	1.25e-02	1.00e+00	1.78e-02
rs17036130	3	12288288	non-coding	1.51e-02	1.27e-02	1.00e+00	1.73e-02
rs13081389	3	12289800	non-coding	4.17e-02	2.98e-02	1.00e+00	1.73e-02
rs1596417	3	12290898	non-coding	4.20e-02	2.99e-02	1.00e+00	1.75e-02
rs13089415	3	12301360	non-coding	2.27e-02	1.62e-02	1.00e+00	1.88e-02
rs6771792	3	12301472	non-coding	2.31e-02	1.63e-02	1.00e+00	1.82e-02
rs4376068	3	185497635	IGF2BP2	7.82e-03	2.16e-03	1.07e-01	1.88e-04
rs6801848	3	185499057	IGF2BP2	1.19e-02	3.81e-03	3.10e-01	9.60e-04
rs4481184	3	185505787	IGF2BP2	3.92e-03	1.12e-03	3.00e-02	5.22e-05
rs11705729	3	185507299	IGF2BP2	3.29e-03	8.83e-04	2.44e-02	4.23e-05
rs11929397	3	185510190	IGF2BP2	5.88e-03	1.22e-03	2.44e-02	4.23e-05
rs7633675	3	185510613	IGF2BP2	6.22e-03	1.28e-03	2.44e-02	4.23e-05
rs16860234	3	185510884	IGF2BP2	1.98e-02	6.43e-03	1.00e+00	1.83e-02
rs4402960	3	185511687	IGF2BP2	3.14e-03	6.87e-04	2.05e-02	3.51e-05
rs16860235	3	185512361	IGF2BP2	3.12e-02	1.33e-02	1.00e+00	4.10e-02
rs7640539	3	185513296	IGF2BP2	4.56e-03	1.07e-03	2.44e-02	4.23e-05
rs7651090	3	185513392	IGF2BP2	5.35e-03	9.58e-04	2.19e-02	3.83e-05
rs6444081	3	185514393	IGF2BP2	4.25e-03	1.01e-03	2.44e-02	4.23e-05
rs7646518	3	185514931	IGF2BP2	4.71e-03	1.09e-03	2.44e-02	4.23e-05
rs7637773	3	185515635	IGF2BP2	4.41e-03	1.04e-03	3.00e-02	6.90e-05
rs4686696	3	185516520	IGF2BP2	4.09e-03	9.86e-04	2.44e-02	4.23e-05
rs6767484	3	185520578	IGF2BP2	3.45e-03	9.03e-04	2.44e-02	4.23e-05
rs7640744	3	185522447	IGF2BP2	2.14e-02	8.37e-03	1.00e+00	1.52e-02
rs11711477	3	185526690	IGF2BP2	5.70e-03	1.18e-03	2.74e-02	5.02e-05
rs1470579	3	185529080	IGF2BP2	5.19e-03	9.31e-04	2.98e-02	5.16e-05
rs6769511	3	185530290	IGF2BP2	6.05e-03	1.25e-03	2.99e-02	5.41e-05
rs9859406	3	185534482	IGF2BP2	6.39e-03	1.31e-03	3.16e-02	5.65e-05
rs2548966	5	134215127	TXNDC15	3.01e-02	1.06e-02	1.00e+00	3.97e-02
rs319602	5	134222164	TXNDC15	2.02e-02	7.07e-03	1.00e+00	3.64e-02

rs319598	5	134240235	PCBD2	3.05e-02	1.08e-02	1.00e+00	4.02e-02
rs319592	5	134252619	PCBD2	3.94e-02	1.21e-02	1.00e+00	4.77e-02
rs319589	5	134255333	PCBD2	2.39e-02	9.74e-03	1.00e+00	3.63e-02
rs6883047	5	134272055	PCBD2	2.35e-02	8.55e-03	1.00e+00	4.71e-02
rs7728823	5	134282777	PCBD2	2.43e-02	9.84e-03	1.00e+00	4.02e-02
rs12658264	5	141764189	non-coding	3.84e-02	6.98e-03	1.00e+00	1.35e-01
rs9348440	6	20641336	CDKAL1	3.19e-02	3.36e-03	1.00e+00	1.79e-03
rs6456364	6	20649254	CDKAL1	4.56e-02	4.08e-03	1.00e+00	1.36e-03
rs9295474	6	20652717	CDKAL1	2.56e-03	2.53e-04	6.90e-04	1.85e-07
rs2328545	6	20653550	CDKAL1	2.93e-02	2.50e-03	1.00e+00	7.99e-04
rs9368216	6	20655110	CDKAL1	4.59e-02	4.15e-03	1.00e+00	8.10e-04
rs4712522	6	20656800	CDKAL1	2.32e-03	2.28e-04	6.11e-04	1.79e-07
rs4712523	6	20657564	CDKAL1	2.40e-03	2.36e-04	7.28e-04	3.31e-07
rs4710940	6	20658012	CDKAL1	5.02e-03	8.62e-04	7.01e-03	4.89e-06
rs6906327	6	20659459	CDKAL1	3.76e-03	3.99e-04	6.65e-03	2.12e-06
rs6456367	6	20659587	CDKAL1	1.48e-03	1.31e-04	5.93e-04	1.73e-07
rs6456368	6	20659806	CDKAL1	1.05e-03	5.92e-05	4.31e-04	1.18e-07
rs6456369	6	20660365	CDKAL1	3.60e-03	3.80e-04	6.52e-03	4.56e-06
rs10946398	6	20661034	CDKAL1	2.02e-03	1.13e-04	8.02e-04	2.40e-07
rs7774594	6	20661143	CDKAL1	1.82e-03	1.70e-04	5.91e-04	1.72e-07
rs7754840	6	20661250	CDKAL1	2.09e-03	2.11e-04	1.23e-03	3.70e-07
rs9460544	6	20661529	CDKAL1	2.63e-03	2.60e-04	5.91e-04	1.73e-07
rs9460545	6	20661550	CDKAL1	1.89e-03	1.78e-04	5.91e-04	1.73e-07
rs4712525	6	20662966	CDKAL1	2.48e-03	2.45e-04	6.04e-04	1.74e-07
rs4712526	6	20663035	CDKAL1	1.40e-03	1.22e-04	6.09e-04	1.74e-07
rs9460546	6	20663632	CDKAL1	2.17e-03	2.19e-04	1.23e-03	3.69e-07
rs742642	6	20665081	CDKAL1	2.97e-02	2.55e-03	1.00e+00	7.52e-04
rs7748382	6	20665549	CDKAL1	1.55e-03	1.39e-04	5.91e-04	1.72e-07
rs7772603	6	20665946	CDKAL1	1.76e-03	1.63e-04	5.53e-04	1.58e-07
rs7752780	6	20666022	CDKAL1	1.62e-03	1.47e-04	5.21e-04	1.48e-07
rs7752906	6	20666055	CDKAL1	1.69e-03	1.55e-04	5.13e-04	1.46e-07
rs9358356	6	20667382	CDKAL1	7.29e-04	3.94e-05	4.74e-04	1.32e-07
rs9356743	6	20667688	CDKAL1	1.69e-02	1.99e-03	6.57e-01	2.24e-04
rs9368219	6	20674691	CDKAL1	1.95e-03	4.83e-05	5.23e-05	1.05e-09
rs1012635	6	20675295	CDKAL1	4.27e-02	4.65e-03	8.35e-01	1.98e-03
rs1569699	6	20679310	CDKAL1	4.99e-04	1.55e-05	3.32e-07	1.36e-11
rs7756992	6	20679709	CDKAL1	1.14e-04	4.37e-07	1.05e-08	0.00e+00
rs9350271	6	20683164	CDKAL1	5.42e-04	1.77e-05	7.74e-07	2.58e-11
rs9356744	6	20685486	CDKAL1	5.83e-04	1.98e-05	7.44e-07	2.58e-11
rs7766070	6	20686573	CDKAL1	2.83e-05	1.43e-07	9.06e-09	0.00e+00
rs9368222	6	20686996	CDKAL1	3.81e-05	2.02e-07	9.06e-09	0.00e+00
rs10440833	6	20688121	CDKAL1	1.60e-05	8.06e-08	9.06e-09	0.00e+00
rs2206734	6	20694884	CDKAL1	8.97e-04	3.05e-05	2.44e-05	7.56e-10
rs6931514	6	20703952	CDKAL1	9.05e-05	3.12e-07	9.06e-09	0.00e+00
rs11753081	6	20705590	CDKAL1	6.76e-04	2.20e-05	1.90e-05	4.95e-10
rs1040558	6	20713706	CDKAL1	7.88e-04	2.49e-05	1.66e-05	4.32e-10
rs9295478	6	20716253	CDKAL1	6.95e-03	6.19e-04	1.29e-03	3.00e-07
rs2328548	6	20716958	CDKAL1	6.28e-04	1.32e-05	1.42e-05	3.78e-10
rs6935599	6	20717095	CDKAL1	9.48e-04	3.31e-05	1.42e-05	3.75e-10
rs9465871	6	20717255	CDKAL1	2.63e-04	1.74e-06	6.30e-06	1.25e-10

rs10946403	6	20717404	CDKAL1	8.44e-04	2.78e-05	1.32e-05	3.42e-10
rs2328549	6	20718240	CDKAL1	3.42e-02	2.81e-03	2.02e-01	8.60e-05
rs9358357	6	20719145	CDKAL1	9.98e-04	3.56e-05	1.42e-05	3.73e-10
rs9368224	6	20719232	CDKAL1	3.12e-04	4.08e-06	1.42e-05	3.73e-10
rs9358358	6	20719393	CDKAL1	1.73e-02	2.03e-03	1.23e-01	5.35e-05
rs9460550	6	20719561	CDKAL1	3.59e-04	6.28e-06	1.42e-05	3.73e-10
rs9356746	6	20720279	CDKAL1	1.22e-02	1.84e-03	1.08e-01	4.81e-05
rs9368226	6	20723057	CDKAL1	4.54e-04	1.08e-05	5.48e-05	1.07e-09
rs12111351	6	20724558	CDKAL1	6.75e-03	6.03e-04	3.17e-03	7.70e-07
rs9356747	6	20725007	CDKAL1	7.14e-03	6.36e-04	3.18e-03	7.72e-07
rs9356748	6	20725097	CDKAL1	1.26e-02	1.88e-03	9.40e-02	4.37e-05
rs7767391	6	20725240	CDKAL1	4.07e-04	8.58e-06	6.90e-05	1.39e-09
rs7747752	6	20725423	CDKAL1	5.52e-03	3.43e-04	2.24e-03	5.65e-07
rs9270986	6	32574060	non-coding	3.16e-02	4.82e-03	1.00e+00	2.29e-03
rs9492055	6	129048640	non-coding	4.73e-02	1.51e-02	1.00e+00	4.80e-02
rs11154899	6	137293890	non-coding	4.89e-02	1.11e-02	1.00e+00	1.09e-01
rs10872465	6	137294656	non-coding	4.86e-02	1.10e-02	1.00e+00	1.09e-01
rs2876354	6	137295352	non-coding	4.96e-02	1.13e-02	1.00e+00	1.09e-01
rs11154900	6	137296161	non-coding	4.93e-02	1.12e-02	1.00e+00	1.10e-01
rs6906007	6	137296300	non-coding	3.31e-02	8.27e-03	1.00e+00	8.99e-02
rs10457653	6	137296895	non-coding	4.83e-02	1.09e-02	1.00e+00	1.09e-01
rs10872466	6	137297967	non-coding	3.91e-02	7.35e-03	1.00e+00	9.89e-02
rs4407733	6	137299152	non-coding	4.99e-02	1.14e-02	1.00e+00	1.09e-01
rs947733	6	137304427	non-coding	3.08e-02	8.46e-03	1.00e+00	6.22e-02
rs849133	7	28192280	JAZF1	4.24e-02	1.58e-02	1.00e+00	1.26e-03
rs849134	7	28196222	JAZF1	2.10e-02	7.80e-03	9.84e-01	1.16e-03
rs849135	7	28196413	JAZF1	3.23e-02	1.16e-02	9.75e-01	1.14e-03
rs10281305	7	54890409	non-coding	4.04e-02	1.18e-02	1.00e+00	1.09e-01
rs4493865	7	54898402	non-coding	4.63e-02	1.52e-02	1.00e+00	1.14e-01
rs2442982	8	20590386	non-coding	3.46e-02	2.29e-02	1.00e+00	1.34e-01
rs4734295	8	96000919	non-coding	9.25e-03	3.49e-03	1.00e+00	2.34e-02
rs10113282	8	96038252	C8orf38	3.87e-02	1.02e-02	1.00e+00	4.08e-02
rs1892012	9	19979945	non-coding	4.46e-02	9.54e-03	1.00e+00	6.54e-02
rs10117648	9	19981497	non-coding	4.49e-02	1.03e-02	1.00e+00	6.93e-02
rs7868773	9	19985150	non-coding	4.53e-02	1.04e-02	1.00e+00	7.03e-02
rs10122799	9	19987293	non-coding	4.11e-02	9.64e-03	1.00e+00	9.47e-02
rs10964378	9	19994736	non-coding	4.69e-02	1.37e-02	1.00e+00	1.09e-01
rs10964380	9	19999413	non-coding	4.66e-02	1.72e-02	1.00e+00	1.09e-01
rs7020996	9	22129579	non-coding	1.25e-03	2.02e-04	1.60e-02	3.66e-06
rs2383208	9	22132076	non-coding	2.24e-03	8.61e-05	3.29e-02	3.61e-06
rs10965250	9	22133284	non-coding	1.32e-03	7.13e-05	5.86e-03	7.57e-07
rs10811661	9	22134094	non-coding	1.17e-03	6.51e-05	8.83e-03	7.70e-07
rs1333051	9	22136489	non-coding	2.71e-03	2.70e-04	7.52e-02	3.98e-05
rs2798253	10	94202905	non-coding	1.60e-02	3.19e-03	9.38e-03	9.56e-06
rs6583813	10	94209939	non-coding	3.57e-02	3.13e-03	1.73e-02	1.22e-05
rs11187007	10	94214580	IDE	2.78e-02	2.75e-03	1.26e-02	4.84e-06
rs2149632	10	94232247	IDE	1.94e-02	2.12e-03	1.80e-02	7.72e-06
rs11187033	10	94262359	IDE	1.89e-02	2.07e-03	1.86e-02	7.07e-06
rs10509645	10	94277866	IDE	3.97e-02	3.02e-03	2.55e-02	1.06e-05
rs2421941	10	94345909	non-coding	4.30e-02	1.38e-03	1.00e+00	9.64e-05

rs10786050	10	94367230	KIF11	3.35e-02	7.05e-04	1.00e+00	1.05e-04
rs10882091	10	94374377	KIF11	3.38e-02	7.22e-04	1.00e+00	9.01e-05
rs10882094	10	94387676	KIF11	2.51e-02	6.69e-04	1.00e+00	8.52e-05
rs10882095	10	94394402	KIF11	4.01e-02	3.08e-03	3.16e-02	1.22e-05
rs10736069	10	94395393	KIF11	2.47e-02	6.52e-04	1.00e+00	7.55e-05
rs7900689	10	94395748	KIF11	2.23e-02	4.37e-04	1.00e+00	7.55e-05
rs6583830	10	94398118	KIF11	2.19e-02	4.18e-04	1.00e+00	7.32e-05
rs10882096	10	94401386	KIF11	3.49e-02	3.25e-03	2.74e-02	1.02e-05
rs11187114	10	94406237	KIF11	3.53e-02	3.30e-03	2.44e-02	8.93e-06
rs4933734	10	94414567	KIF11	1.11e-02	2.96e-04	1.00e+00	1.55e-05
rs7911264	10	94436851	non-coding	1.15e-02	3.11e-04	8.34e-02	4.18e-07
rs2488087	10	94446041	non-coding	1.07e-02	1.04e-04	8.34e-02	4.18e-07
rs10882100	10	94460687	non-coding	1.04e-02	9.52e-05	8.65e-02	4.28e-07
rs1111875	10	94462882	non-coding	3.02e-03	3.61e-04	1.24e-03	4.82e-07
rs12778642	10	94464307	non-coding	2.90e-03	2.82e-04	9.05e-04	4.49e-07
rs5015480	10	94465559	non-coding	1.10e-03	7.74e-05	8.78e-04	1.12e-07
rs10882102	10	94466495	non-coding	2.80e-03	3.25e-04	1.30e-03	4.77e-07
rs11187144	10	94469980	non-coding	8.09e-03	1.70e-03	1.03e-02	8.93e-06
rs7087591	10	94473629	non-coding	8.35e-03	1.73e-03	9.41e-03	6.14e-06
rs10748582	10	94477219	non-coding	7.35e-03	1.35e-03	6.65e-03	3.33e-06
rs7923837	10	94481917	non-coding	8.61e-03	1.77e-03	8.97e-03	4.89e-06
rs7923866	10	94482076	non-coding	7.56e-03	1.44e-03	8.37e-03	4.89e-06
rs7917983	10	114732882	TCF7L2	6.55e-03	4.35e-05	1.78e-01	5.65e-07
rs7901275	10	114732906	TCF7L2	4.86e-03	5.33e-05	1.63e-01	3.70e-07
rs4074720	10	114748497	TCF7L2	1.98e-04	9.75e-09	7.42e-09	0.00e+00
rs4074718	10	114748617	TCF7L2	7.27e-05	5.26e-10	6.32e-09	0.00e+00
rs17747324	10	114752503	TCF7L2	6.05e-08	4.00e-13	0.00e+00	0.00e+00
rs7901695	10	114754088	TCF7L2	5.21e-09	8.64e-15	0.00e+00	0.00e+00
rs4506565	10	114756041	TCF7L2	1.07e-10	4.28e-20	0.00e+00	0.00e+00
rs7903146	10	114758349	TCF7L2	2.40e-11	4.61e-22	0.00e+00	0.00e+00
rs10885402	10	114761697	TCF7L2	4.88e-05	3.72e-10	7.42e-09	0.00e+00
rs6585198	10	114762237	TCF7L2	7.86e-05	6.60e-10	7.42e-09	0.00e+00
rs4132670	10	114767771	TCF7L2	3.25e-09	5.32e-15	0.00e+00	0.00e+00
rs6585200	10	114768609	TCF7L2	9.65e-05	9.05e-10	8.10e-09	0.00e+00
rs6585201	10	114768783	TCF7L2	8.37e-05	7.78e-10	8.10e-09	0.00e+00
rs7904519	10	114773927	TCF7L2	1.89e-04	1.68e-09	9.06e-09	0.00e+00
rs10885405	10	114777670	TCF7L2	1.26e-04	1.31e-09	9.06e-09	0.00e+00
rs10885406	10	114777724	TCF7L2	1.32e-04	1.43e-09	9.06e-09	0.00e+00
rs10787472	10	114781297	TCF7L2	1.20e-04	1.18e-09	1.00e-08	0.00e+00
rs7924080	10	114787012	TCF7L2	1.08e-04	1.04e-09	9.06e-09	0.00e+00
rs12243326	10	114788815	TCF7L2	9.89e-09	4.46e-14	0.00e+00	0.00e+00
rs7077039	10	114789077	TCF7L2	1.02e-04	1.19e-10	8.32e-09	0.00e+00
rs7900150	10	114793823	TCF7L2	6.60e-05	8.16e-11	7.42e-09	0.00e+00
rs7100927	10	114796048	TCF7L2	1.79e-04	2.12e-10	7.42e-09	0.00e+00
rs7895340	10	114801525	TCF7L2	5.81e-05	4.77e-11	7.42e-09	0.00e+00
rs11196200	10	114801938	TCF7L2	1.41e-04	1.62e-10	7.42e-09	0.00e+00
rs11196205	10	114807047	TCF7L2	2.08e-04	8.02e-09	7.00e-08	0.00e+00
rs12255372	10	114808902	TCF7L2	2.25e-07	4.11e-12	0.00e+00	0.00e+00
rs12265291	10	114810240	TCF7L2	1.60e-04	4.87e-09	6.65e-08	0.00e+00
rs11196208	10	114811316	TCF7L2	1.51e-04	3.35e-09	6.65e-08	0.00e+00

rs7077247	10	114812071	TCF7L2	1.69e-04	6.27e-09	7.00e-08	0.00e+00
rs12718338	10	114813047	TCF7L2	2.32e-04	1.35e-08	1.11e-07	0.00e+00
rs10832778	11	17394073	B7H6	2.82e-02	1.64e-02	1.00e+00	1.53e-01
rs1557765	11	17403639	non-coding	9.61e-03	6.34e-03	1.00e+00	3.68e-02
rs5215	11	17408630	KCNJ11	8.91e-03	4.50e-03	1.00e+00	2.36e-02
rs7124355	11	17412960	non-coding	2.06e-02	1.57e-02	1.00e+00	4.76e-02
rs757110	11	17418477	ABCC8	9.98e-03	6.16e-03	1.00e+00	2.67e-02
rs1877527	12	71405206	non-coding	4.07e-02	8.08e-03	1.00e+00	3.68e-02
rs11178531	12	71408690	non-coding	2.55e-02	5.63e-03	1.00e+00	2.36e-02
rs7132840	12	71411561	non-coding	3.81e-02	9.35e-03	1.00e+00	2.19e-02
rs2063591	12	71411855	non-coding	2.59e-02	5.72e-03	1.00e+00	3.59e-02
rs7957932	12	71421552	non-coding	4.76e-02	1.31e-02	1.00e+00	4.31e-02
rs1512991	12	71422768	non-coding	1.77e-02	4.99e-03	1.00e+00	2.33e-02
rs7956274	12	71424402	non-coding	1.81e-02	5.08e-03	1.00e+00	2.36e-02
rs7959965	12	71425164	non-coding	1.85e-02	5.17e-03	1.00e+00	2.36e-02
rs7298255	12	71428069	non-coding	1.30e-02	4.01e-03	1.00e+00	2.07e-02
rs10784891	12	71429798	non-coding	1.64e-02	4.73e-03	1.00e+00	2.79e-02
rs7955901	12	71433293	non-coding	2.75e-02	6.07e-03	1.00e+00	3.17e-02
rs4760894	12	71438923	non-coding	2.90e-02	7.53e-03	1.00e+00	3.64e-02
rs4760785	12	71438945	non-coding	2.63e-02	5.81e-03	1.00e+00	3.69e-02
rs4760895	12	71439127	non-coding	2.67e-02	5.90e-03	1.00e+00	3.77e-02
rs7138300	12	71439589	non-coding	2.71e-02	5.99e-03	1.00e+00	3.86e-02
rs1913201	12	71439825	non-coding	2.86e-02	7.44e-03	1.00e+00	3.86e-02
rs10879240	12	71443285	non-coding	3.27e-02	8.85e-03	1.00e+00	3.88e-02
rs7313973	12	71444058	non-coding	1.56e-02	4.58e-03	1.00e+00	4.02e-02
rs1554522	17	25913172	KSR1	4.36e-02	1.45e-02	1.00e+00	2.13e-01

6. Details of simulations results in Section 5.1. Figure 4 shows the false discovery proportion (FDP) in a replicability analysis (top), and in an analysis to discover associations (bottom). The variation in FDP decreases with M, and is very small for $M = 10^5$.

References.

Benjamini, Y., Heller, R. and Yekutieli, D. (2009). Selective Inference in Complex Research. *Philosophical Transactions of the Royal Society A* **267** 1–17.

Efron, B. and Tibshirani, R. (2002). Empirical Bayes Methods and False Discovery Rates for Microarrays. *Genetic Epidemiology* **23** 70–86.

Storey, J. (2002). A Direct Approach to False Discovery Rates. *Journal of the Royal Statistical Society, Series B* **64 (3)** 79–498.

Department of statistics and Operations research
Tel-Aviv University
Tel-Aviv, Israel
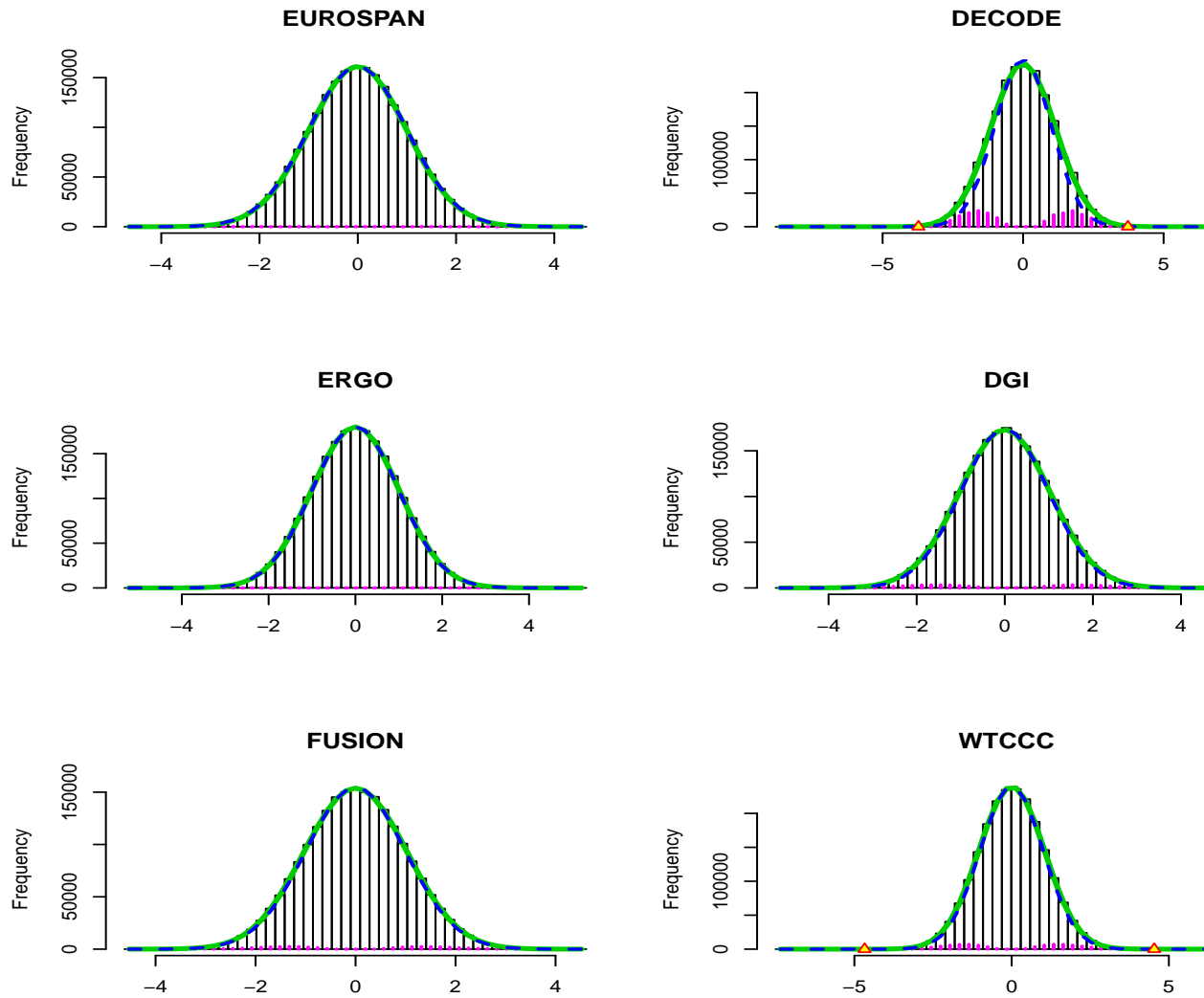E-mail: ruheller@math.tau.ac.il
        yekutiel@post.tau.ac.il

FIG 3. *The histogram of z-scores for each of the six T2D GWA studies. The heavy curve is the estimate $\hat{f}_i(z)$ for the mixture density $f(z)$, scaled to match the histogram area. Dashed curve is scaled estimate $\hat{\pi}_0(i)f_0(z)$, where $f_0(z)$ is the standard normal density. The estimated non-null counts are shown in pink.*
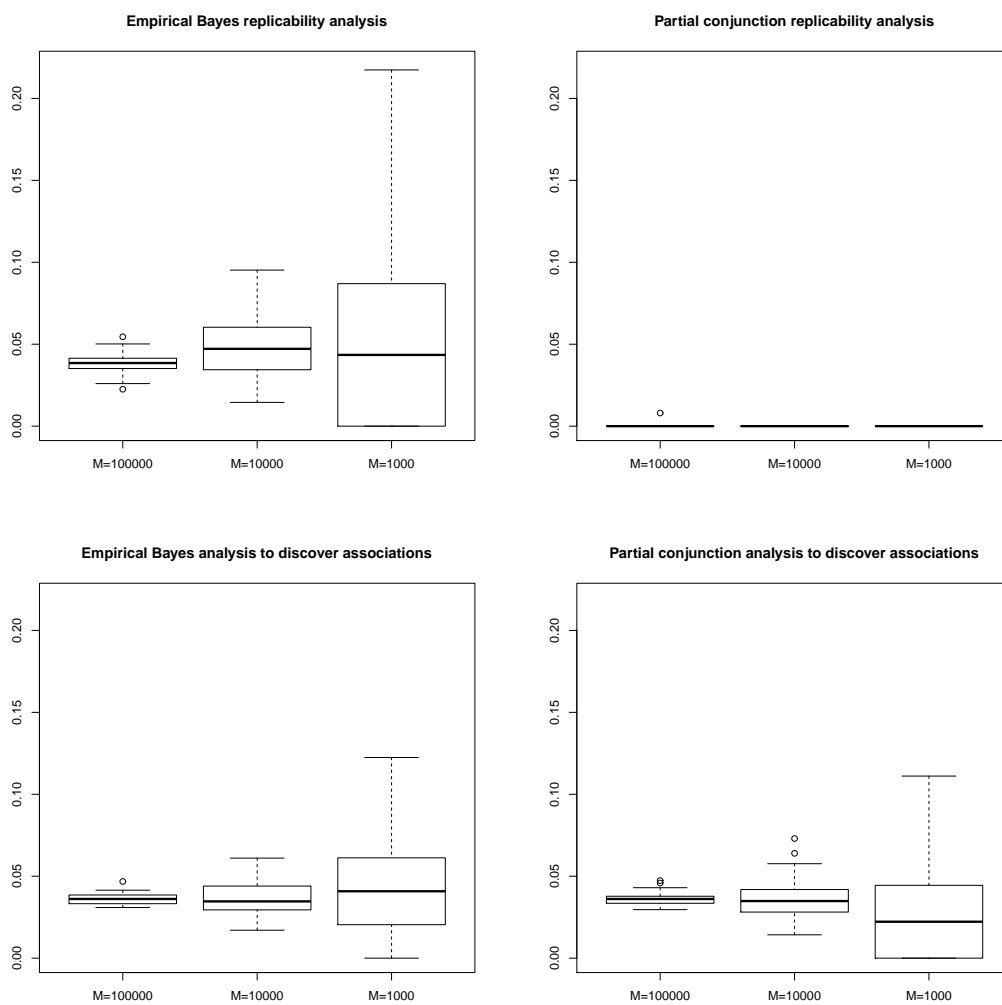
FIG 4. *Replicability analysis (top) and analysis to discover association (bottom): Boxplots of FDP for M=100000, M=10000, and M=1000 for empirical Bayes analysis (left) and the analysis of BHY09 (right).*