Published in final edited form as:
J Bioinform Comput Biol. 2014 August ; 12(4): 1450018. doi:10.1142/S0219720014500188.

COMPARATIVE ANALYSIS OF FALSE DISCOVERY RATE METHODS IN CONSTRUCTING METABOLIC ASSOCIATION NETWORKS

IMHOI KOO[1], **SEN YAO**[2], **XIANG ZHANG**[3], and **SEONGHO KIM**[4,*]

[1]Department of Chemistry, University of Louisville, Louisville, Kentucky 40292, USA

[2]Department of Chemistry, University of Louisville, Louisville, Kentucky 40292, USA

[3]Department of Chemistry, University of Louisville, Louisville, Kentucky 40292, USA

[4]Biostatistics Core, Karmanos Cancer Institute, Department of Oncology, Wayne State University Detroit, Michigan 48201

Abstract

Gaussian graphical model (GGM)-based method, a key approach to reverse engineering biological networks, uses partial correlation to measure conditional dependence between two variables by controlling the contribution from other variables. After estimating partial correlation coefficients, one of the most critical processes in network construction is to control the false discovery rate (FDR) to assess the significant associations among variables. Various FDR methods have been proposed mainly for biomarker discovery, but it still remains unclear which FDR method performs better for network construction. Furthermore, there is no study to see the effect of the network structure on network construction. We selected the six FDR methods, the linear step-up procedure (BH95), the adaptive linear step-up procedure (BH00), Efron's local FDR (LFDR), Benjamini-Yekutieli's step-up procedure (BY01), Storey's q-value procedure (Storey01), and Storey-Taylor-Siegmund's adaptive step-up procedure (STS04), to evaluate their performances on network construction. We further considered two network structures, random and scale-free networks, to investigate their influence on network construction. Both simulated data and real experimental data suggest that STS04 provides the highest true positive rate or F1 score, while BY01 has the highest positive predictive value in network construction. In addition, no significant effect of the network structure is found on FDR methods.

Keywords

false discovery rate; Gaussian graphical model; network construction; random network; scale-free network

[1]imhoi.koo@louisville.edu. [2]sen.yao@louisville.edu. [3]xiang.zhang@louisville.edu. [*] Corresponding author. [4]kimse@karmanos.org.

1. Background

Research interests in bioinformatics have been moved from individual target, such as gene, protein, and metabolite, into large-scale systems, called '-omics', to understand the behavior and the association of an entire cellular system. Metabolomics is the latest of the 'omics' sciences, and has emerged as a powerful method for a systematic study of small-molecules, called 'metabolites'. An active research topic in metabolomics is to look for associations among metabolites at the system level to obtain physiological snapshot of cellular status, e.g., Kyoto Encyclopedia of Genes and Genomes (KEGG) [1] and Integrated Pathway Gene Relationship Database (IntPath) [2]. Network analysis plays a key role in understanding the biological systems.

One of the widely used approaches to biological network construction is Gaussian graphical model (GGM) [3]. GGM represents association between two variables by a partial correlation coefficient which measures conditional dependence or independence between a pair of variables after controlling other variables. After estimating a partial correlation from metabolite data, the significance of each edge is then determined by a statistical test to see if the two variables (nodes) connected by this edge are truly connected. If the partial correlation coefficient between two variables is greater than a certain cut-off value, the edge corresponding to the variables is considered as connected. However, in the context of network reconstruction, the number of edges corresponds to the number of hypotheses and the number of edges is often enormous, so the multiple comparison correction is indispensable. The most used criterion in network construction is the false discovery rate (FDR), which was introduced by Benjamin and Hochberg [4] for analysis of microarray data [5]. The FDR approach has been employed in several studies to reconstruct biological networks [6, 7]. Note that we do not consider a method that uses L1 (lasso) regularization to calculate sparse partial correlation coefficient matrix, such as graphical Lasso [8], in this study, because existing L1 regularization-based approaches are required no FDR methods for network construction, which is beyond the scope of this paper.

To properly control the rate of false discoveries, various FDR methods have been proposed [4, 3]. Although the local FDR method was often used [6, 4, 5], it still remains unknown which FDR method performs better for the network reconstruction. In addition, it is known that the structure of biological metabolite networks resemble the scale-free network [6]. However, the effect of the network structure on FDR remains unclear.

We evaluated the performance of the six FDR methods, including the linear step-up procedure [4], the adaptive linear step-up procedure [9], local FDR [7], Benjamini-Yekutieli's step-up procedure [2], Storey's q-value procedure [1], and Storey-Taylor-Siegmund's adaptive step-up procedure [3], to reconstruct metabolite association network using simulation data and real biological experimental data. The linear step-up and Storey's q-value procedures were chosen since they are widely used approaches, while the local FDR was considered based on several applications to network construction [6, 4, 5]. Kim and de Wiel [8] studied the effect of dependence among variables on FDR control procedures and concluded that the adaptive linear step-up procedure is the most optimal under dependence.

For this reason, we included the adaptive linear step-up FDR procedure. The other two FDR procedures were considered because of their robustness to the dependency of variables.

Furthermore, to investigate the effect of the network structure through simulation data, we considered two types of network structures, such as random and scale-free networks for true network. The random network [9] has a degree which follows a binomial distribution, while the scale-free network [20], of which the degree is dominated by power distribution, is well fitted with some real life network structures such as metabolite network [6, 1], social network [2], and World Wide Web [3]. Takahashi et al. [4] have recently concluded that scale-free networks best describe protein-protein interaction networks of different species and Attention Deficit Hyperactivity Disorder (ADHD) based on statistical testing.

The rest of this paper is organized as follows. Section 2 presents a brief description of the six FDR methods and surveys R/Bioconductor packages currently available for FDR calculation. In Section 3, the six FDR methods were evaluated by applying to simulated data as well as real metabolite data. Section 4 concludes with some discussion. Supplementary Information is given in a complementary document for Figures S1-S8 and Tables S1-S4.

2. Methods and Materials

2.1. Network construction

Let $X = (x_{ij}) \in \mathbb{R}^{n\times p}$ be a data matrix with p variables and n sample size, $\mathbf{x}_i$ a column vector as $(x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$, and the sample mean vector $\bar{\mathrm{x}} = \frac{1}{n}\Sigma_{i=1}^{n}\mathrm{x}_i$, where x_{ij} is the ith observation on the jth random variable and $p < n$. The sample covariance is $p \times p$ matrix defined by

$$S = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathrm{x}_i - \bar{\mathrm{x}}\right)\left(\mathrm{x}_i - \bar{\mathrm{x}}\right)^{\mathrm{T}} \quad (1)$$

Then a partial correlation coefficient ρ_{ij}, $i,j = 1, \ldots, p$, is defined from an inverse variance-covariance matrix $S^{-1} = (\theta_{ij})$ as follows [3]:

$$\rho_{ij} = -\frac{\theta_{ij}}{\sqrt{\theta_{ii}\theta_{jj}}}, \quad i \neq j \quad (2)$$

Once all the partial correlation coefficients are estimated, statistical testing should be carried out over each of the estimated partial correlation coefficients to find significant associations. Then, the association network is constructed using the edges with the significant partial correlation coefficients. However, the statistical procedure to identify significant partial correlations involves a multiple comparison problem since many hypotheses are tested simultaneously. To overcome this difficulty, false discovery rate (FDR) control is used in network construction.

2.2. False discovery rate (FDR)

Consider m hypotheses to be tested simultaneously,

$$H_{0i} \quad \text{versus} \quad H_{1i}, i=1,\ldots,m, \quad (3)$$

where H_{0i} and H_{1i} are a null and an alternative hypotheses, respectively, for the *i*th partial correlation coefficient. Suppose the number of true null hypotheses is m_0 and then the number of true alternative hypotheses is $m - m_0$. Among the m_0 null hypotheses, V hypotheses are declared alternative, which are false positives, and U hypotheses are declared null, which are true negatives. As for the $m - m_0$ alternative hypotheses, the tests declare that T hypotheses are null, which are false negatives, and S hypotheses are declared alternative, which are true positives. As a result, R (= $S + V$) hypotheses are declared alternative and $m - R$ (= $U + V$) hypotheses are declared null. Table 1 summarizes these hypothesis tests in a conventional form. Then the false discovery rate (FDR) is defined as

$$FDR=E[Q]=E\left[\frac{V}{V+S}\right]=E\left[\frac{V}{R}\right] \quad (4)$$

where Q, an unobserved random variable, is the proportion of the rejected null hypotheses.

Various procedures have been proposed to control the false discovery rate for the multiplicity problem [5]. In this study, we focus on the following six FDR methods to compare their performances on network construction:

2.2.1. The linear step-up procedure (BH95)—The linear step-up procedure was proposed by Benjamini and Hochberg under the assumption that all test statistics are independent [4]. Let $p_1 \le p_2 \le \ldots \le p_m$ be an order of p-values corresponding to null hypotheses $H_{01}, H_{02}, \ldots, H_{0m}$. To control FDR at q, one reject all H_{0i}, $i = 1, 2, \ldots, k$, where k is determined as follows:

$$k=max\{i|p_i \le q \cdot i/m\} \quad (5)$$

2.2.2. The adaptive linear step-up procedure (BH00)—Benjamini and Hochberg in 2000 proposed an adaptive linear step-up approach to FDR calculation [9], which is more powerful than the linear step-up procedure (BH95). To control FDR at q, one rejects k hypotheses H_{0i}, $i = 1, 2, \ldots, k$, corresponding to

$$k=max\{i|p_i \le q \cdot i/\hat{m}_0\} \quad (6)$$

where $p_1 \le p_2 \le \ldots \le p_m$ are the ordered p-values. The estimate of m_0 is calculated by

$$\hat{m}_0=min\left\{\left(\frac{1}{S_j}+1\right), m\right\} \quad (7)$$

where $S = (1 - p_j)/(m + 1 - j)$ with $j = \min\{i|S_i < S_{i-1}\}$.

2.2.3. Efron's local FDR (LFDR)—Define the prior probabilities of null and alternative hypotheses as p_0 and $p_1 = 1 - p_0$, repecitvely, with corresponding densities $f_0(z)$ and $f_1(z)$ for the test statistic Z, where

$$p_0 = Pr\{null \quad is \quad true\}, \quad (8)$$

$$p_1 = Pr\{alternative \quad is \quad true\}, \quad (9)$$

$$f_0(z) \quad \text{is the density of} \quad Z \quad \text{if null is true,} \quad (10)$$

and

$$f_1(z) \quad \text{is the density of} \quad Z \quad \text{if alternative is true.} \quad (11)$$

Then the mixture density of statistic *Z* is defined by

$$f(z) = p_0 f_0(z) + p_1 f_1(z). \quad (12)$$

By Bayes' theorem, the local false discovery rate is defined by [10]:

$$fdr(z) \equiv Pr\{null \quad is \quad true|z\} = p_0 f_0(z) / f(z). \quad (13)$$

2.2.4. Benjamini-Yekutieli's step-up procedure (BY01)—Benjamini and Yekutieli [2] demonstrated that the linear step-up procedure (BH95) can control the FDR even for positive regression dependence structures and further proposed another version of BH95 to control FDR under arbitrary dependence structures. Namely, under dependency on each of the test statistics, Benjamini and Yekutieli [2] showed that when the adaptive linear step-up procedure (BH00) is conducted with $q/\left(\Sigma_{i=1}^{m}\frac{1}{i}\right)$ instead of *q* in the Eq. (5), it always controls the FDR at the level less than or equal to $\frac{m_0}{m}q$.

2.2.5. Storey's q-value procedure (Storey01)—The FDR is defined as follows by Benjamini and Hochberg [4]:

$$FDR = Pr(R>0)\,\mathrm{E}[V/R|R>0]. \quad (14)$$

However, the FDR in Eq. (14) equals to one, if $m = m_0$, with any critical value when $Pr(R = 0) = 0$. To avoid this issue, Storey [1] employed a modified version of the FDR, positive false discovery rate (pFDR), in controlling FDR rate. pFDR is defined as

$$pFDR = \mathrm{E}[V/R|R>0]. \quad (15)$$

Denote the *m* p-values by $p_1, p_2, \ldots, p_m$. Then the FDR when rejecting all null hypotheses with $p_i \le s$, given a cut-off value *s*, $0 < s \le 1$, is estimated by

$$\mathrm{p}\hat{FDR}_\lambda(s) = \frac{\hat{\pi}_0(\lambda)\cdot m\cdot s}{\#\{p_i \le s\}}, \quad (16)$$

where $\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda; i=1,\ldots,m\}}{m(1-\lambda)}$ with a tuning parameter λ ∈ [0,1). Under independent condition, we can estimate the q-value of the *i*th p-value, p_i, by pFDR(s) defined in Eq. (16):

$$q - \text{value} \quad (p_i) = inf_{s \geq p_i} \quad pF\hat{D}R_\lambda(s). \quad (177)$$

We refer to the results of Storey [1, 6] and Storey and Tibshirani [7] for detailed information.

2.2.6. Storey-Taylor-Siegmund's adaptive step-up procedure (STS04)—Storey et al. [3] introduced an extension of the conservative bias result of [6], by assuming that the p-values of the true null hypotheses are independent and uniformly distributed. Their FDR controlling step-up method has two minor modifications to pFDR defined in Eq. (16). The estimate of $\hat{\pi}_0(\lambda)$ was modified to guarantee that $\hat{\pi}_0(\lambda) > 0$ if λ > 0. The modified estimated is, for λ > 0, defined by

$$\hat{\pi}_0^*(\lambda) = \frac{\#\{p_i > \lambda; i=1,\ldots,m\} + 1}{(1-\lambda)\cdot m} \quad (188)$$

where $\hat{\pi}_0^*(\lambda=0) = \hat{\pi}_0(\lambda=0) = 1$. Then, by limiting the cut-off value *s* only to the interval [0, λ], the modified estimate of pFDR(s) is obtained by

$$pF\hat{D}R_\lambda^*(s) = \begin{cases} \frac{\hat{\pi}_0^*(\lambda)\cdot m\cdot s}{\#\{p_i \leq s\}} & if \quad s \leq \lambda \\ 1 & if \quad s > \lambda \end{cases} \quad (19)$$

where λ ∈ [0,1) is a tuning parameter. The q-value can further be calculated by replacing $pF\hat{D}R_\lambda(s)$ of Eq. (17) with $pF\hat{D}R_\lambda^*(s)$ of Eq. (19),

2.3. R/Bioconductor packages

A number of R/Bioconductor packages have been introduced for controlling FDR. Supplementary Information Table S1 lists various FDR packages which are available in R (CRAN) and Bioconductor projects (www.r-project.org and www.bioconductor.org, respectively).

Most packages require p-values as input, except for the R packages *locfdr* and *fdrtool*. The *locfdr* uses z-scores (i.e., z-statistics) [10], as well as *fdrtool* allows multiple scores, such as z-scores, correlation coefficients, and t-scores (i.e., t-statistics), as input. Interestingly, the R package *stats*, which is one of the basic R packages, provides a function 'p.adjust' to control FDR. The function 'p.adjust' provides users with two FDR methods, the linear step-up procedure (BH95) and Benjamini-Yekutieli's step-up procedure (BY01). It also includes the Bonferroni correction and several less conservative corrections such as the methods developed by Holm [8], Hochberg [9], and Hommel [30].

The function *fdrtool* implemented in the contributed R package *fdrtool* is one of the widely used functions for network construction. It includes both tail area-based FDR, which is called frequentist FDR, and local FDR with various input types, such as p-value, z-scores,

correlation coefficients, and t-scores. In addition, it provides users with graphical outputs by visualizing the performance of the FDR control. Typical graphical outputs are the histogram and the density of the fitted two-component model, the corresponding cumulative density functions, and the dependence of the test statistics between the local FDR and the tail area-based FDR [1].

Recently, an R package *mutoss* has been developed for multiple hypotheses testing with a graphical user interface (GUI) through its GUI version, *mutossGUI*. The R package *mutoss* requires the R package *fdrtool* and the Bioconductor packages *multtest* and *qvalue*, including the local FDR, the linear step-up procedure (BH95), the adaptive linear step-up procedure (BH00), and Benjamini-Yekutieli's step-up procedure (BY01). In addition to these FDR methods, it also provides users with a variety of FDR methods, such as Storey-Taylor-Siegmund adaptive step-up procedure [3], Benjamini-Liu step-down procedure [2], and Blanchard-Roquain step-up procedure [3], including various non-FDR methods for multiple hypothesis testing. The readers should consult the R/Bioconductor websites for more detailed descriptions of these packages including other packages shown in Supplementary Information Table S1.

As discussed in Section 2.2, we employed six FDR methods in this study. Although we found that the R package *mutoss* has all of the six FDR methods, other packages also have these six FDR methods. Therefore, we validated these packages to see if all the packages generate the same result for each of the six FDR methods. To do this, we compared the results of the R package *mutoss* with these of other R/Bioconductor packages, such as *fdrtool*, *stats* ('p.adjust'), *locfdr*, *multtest*, and *qvalue*, and found no difference among these R packages (data not shown). Thus, we used the R package *mutoss* in this study for the following six FDR methods: the linear step-up procedure (BH95) [4], the adaptive linear step-up procedure (BH00) [9], Efron's local FDR (LFDR) [10], Benjamini-Yekutieli's step-up procedure (BY01) [2], Storey's q-value procedure (Storey01) [1], and Storey-Taylor-Siegmund's adaptive step-up procedure (STS04) [3]. These methods are implemented by the input parameters *BH*, *adaptiveBH*, *pval2locfdr*, *BY*, *Qvalue*, and *adaptiveSTS*, respectively, in the function 'mutoss.apply' of the R package *mutoss*.

2.4. Data

2.4.1. Simulated data—The simulated data are generated based on the two true network topologies such as the scale-free network [19] and the random network [18]. The scale-free network is a network whose probability of degree k follows a power-law distribution:

$$P(k) \sim k^{-\gamma} \quad (190)$$

where $P(k)$ is the fraction of nodes in the network having k edges or links and γ is a parameter of the distribution. It is known that the structure of biological networks resembles the scale-free network [6]. On the other hand, the random network is a network that is randomly generated. In fact, the scale-free network is a special case of random networks in the context of network theory since it is generated by a random distribution. However, in this study, we call the network whose probability of degrees follows a power-law

distribution the scale-free network and the network whose probability of degrees follows a binomial distribution the random network.

To generate the true network structure for scale-free, the R package *iGraph* was utilized, while the R package *GeneNet* [4] was employed for the random network construction. As for scale-free network, the R package *iGraph* requires users to set the parameter that is the number of growth of next step and we used 2 for this number in this study.

Besides the true network structure, we considered another criterion, the network density which is a proportion of the truly significant edges, when generating the simulated data sets. Note that the number of samples (observation) and the number of variables are set to $n = 200$ and $p = 100$, respectively. We used the four different network densities such as 0.025, 0.05, 0.1, and 0.25. We then generated 1000 data sets according to two network structures and four network complexities, resulting in a total of eight cases. Note that each of the eight cases has 1000 data sets.

Once the simulated data were generated, we applied the six FDR methods, BH95, BH00, LFDR, BY01, Story01, and STS04, to the eight cases to evaluate their performances on network construction. We further investigated the effect of the cut-off values on network construction using the following six values, 0.05, 0.1, 0.15, 0.2, 0.25, and 0.3, as the cut-off values to control FDR.

2.4.2. Real data—The real biological data consist of 12 samples of metabolites extracted from mouse liver. Metabolites were detected from a linear trap quadruple Fourier transform ion cyclotron resonance mass spectrometer (LTQ-FTICR-MS) using direct infusion. After preprocessing using ‘MetSign’ software [4], we obtained 94 compounds that are present in all of the 12 samples. Since the sample size is less than the number of variables (i.e., 12<94), the conventional GGM could not be used. A modified GGM approach was, therefore, applied to the real data for calculation of the partial correlation coefficients. To do this, we used the partial least squares-based method using the approach introduced by [7].

2.5. Performance criteria

The constructed networks by each of the six FDR methods were evaluated using the following measurements:

1. The number of discovery (Discov): It is the number of outcomes which are predicted as positive.
2. The true positive (TP): It is the number of positive outcomes whose actual values are positive.
3. The true positive rate (TPR): It is the probability that the prediction outcome is positive when the actual value is positive and is defined by

$$TPR=\frac{TP}{TP+FN} \quad (201)$$

4. The false positive rate (FPR): It is the probability that the prediction outcome is positive when the actual value is negative and is defined by

$$FPR=\frac{FP}{FP+TN} \quad (212)$$

5. The positive predictive value (PPV): It is the probability that the actual value is positive when the prediction outcome is positive and is defined by

$$PPV=\frac{TP}{TP+FP} \quad (223)$$

6. F1 score: It is a measure of accuracy, which is the harmonic average of TPR and PPV, and is defined by

$$F1=2\cdot\frac{TPR\cdot PPV}{TPR+PPV} \quad (234)$$

7. Receiver operating characteristic (ROC) curve: It is created by plotting between TPR and FPR according to various cut-off values [5].
8. The area under ROC curve (AUC): It is used to measure the efficiency of a method [6].

3. Results

The six FDR methods were compared by applying each of them to the simulated and real data. The six FDR approaches are the linear step-up procedure (BH95), the adaptive linear step-up procedure (BH00), Efron's local FDR (LFDR), Benjamini-Yekutieli's step-up procedure (BY01), Storey's q-value procedure (Storey01), and Storey-Taylor-Siegmund's adaptive step-up procedure (STS04), as described in Section 2.3.

3.1. Analysis of the simulated data

We first evaluated the performances of each FDR method for the network construction using the simulated data. As stated in Section 2.4, we generated 1000 data sets for each of the cases according to the network structures (scale-free network and random network) and the network complexities (0.025, 0.05, 0.1, and 0.25), resulting in a total of eight cases. Then the six FDR methods were applied to each of the eight cases by varying the cut-off values of the FDR.

Figure 1 displays the value of the true positive rate (TPR) with respect to different FDR methods, network densities, and FDR cut-off values. The results of random networks are in the upper row and the bottom row represents the results of scale-free networks. In case of random networks, the value of TPR increases with the increase of the cut-off value but decreases with the increase of the network complexity. Interestingly, the performances of LFDR and BY01 are clearly distinguished from the other four methods, BH95, BH00, Storey01, and STS04. These four FDR methods perform very similarly regardless of the

density and the cut-off value, while LFDR and BY01 have very comparable TPR at the low density but have different TPR as the network becomes complex. At the density of 2.5%, the TPRs of LFDR and BY01 become either very similar to each other or intertwined at the small cut-off value. To confirm this behavior, we extended the range of the network density and the cut-off value as shown in Supplementary Information Figure S1. The upper row of the figure evidently shows that the TPR of BY01 becomes larger than that of LFDR at the smaller cut-off value with a density of 1%. The similar trend can be observed for the scale-free networks as shown in the bottom rows of both Figure 1 and Supplementary Information Figure S1. Supplementary Information Figure S2 further depicts the number of true positives (TP). Likewise, the TP becomes smaller as the network becomes complex, while the TP becomes larger as the cut-off value becomes larger. Overall, Story01 is the leading FDR method across all densities and cut-off values, followed very closely by STS04, while BY01 shows the worst performance in terms of TPR.

The plots of the predictive positive value (PPV) are shown in Figure 2. Likewise to TPR, there is no apparent influence of the network structure on the performance of PPV and the highest PPV becomes smaller as the network becomes complex although the overall trend of PPV seems affected by the network density. Moreover, LFDR and BY01still show distinct trends from other FDR methods. However, differently from the performance of PPV, the performances of all six methods are completely flipped when the network density is 5%. That is, LDFR and BY01 perform better than the others in relatively small network densities, while their PPVs become smaller than these of the others in large network complexity. More specifically, the PPVs of BY01 and BH95 are larger than these of the others when the network is less complex, while LFDR and Stoery01 perform better than the others when the network becomes more complex. The same trends also hold true for the scale-free networks as can be seen in the bottom row of the Figure 2. We can see the comparable trends even when the network density is either 1% or 50%, as depicted in Supplementary Information Figure S3. The number of discovery (Discov) is further considered as displayed in Supplementary Information Figure S4. The trend of Discov seems very similar to that of TPR or TP, not that of PPV. Namely, LFDR and BY01 always perform worse than the other four methods in terms of Discov across all the densities. Overall, BY01 shows the best performance among the six FDR methods when the network is less complex, but, when the network is more complex, Storey01 performs the best in terms of PPV.

Figure 3 shows the plots of F1 scores with respect to the random and scale-free networks. As stated in Section 2.5, the F1 score is the harmonic mean of TPR and PPV, resulting in a combined trend of PPV and TPR. Therefore, the optimal cut-off value can be obtained by taking the value with the maximal F1 score. In particular, when the network density is more than or equal to 5%, the highest F1 scores achieved by the six FDR methods is less than 50% in case of the random networks, while all the densities report the F1 scores less than 50% across all the densities as for the scale-free networks. Another interesting point is that the F1 score is always monotonically increasing with increasing the cut-off value except for the random network with density of 2.5%. To inspect this, we extended the ranges of the density and the cut-off value. The Supplementary Information Figure S5 depicts the results, suggesting that there is not an optimal cut-off value and the maximal F1 score is less than

50% if the network density exceeds 10%. However, BY01 has appears to have an optimal cut-off at the density of 1% when the structure follows the random network. The optimal cut-off values of the other FDR methods increase as the density becomes more complex, as shown in the first and second columns of Figure S5. In addition, given the same complexity of the network, the optimal cut-off values of the scale-free network are shifted to the right compared to the random network. The detected optimal cut-off value is 0.1~ 0.2 for BY95, BH00, Storey01, and STS04 but it is more than 0.4 in case of LFDR. Overall, BY95, BH00, Storey01, and STS04 perform better than LFDR and BY01 in terms of F1 values and the F1 score of BY01 has the smallest value among the six FDR methods.

We also created the ROC curves for the six FDR methods with different network densities corresponding to several cut-off values, as can be seen in Figure 4. The trends of LFDR and BY01 are definitely distinct from the other four FDR methods. The ROC curves of LFDR and BY01 are located at the left-most bottom position, while these of others are located at the upper position more shifted to the right, regardless of the network structure. This suggests that LFDR and BY01 are less powerful (since their TPRs are lower) but more conservative (since their FPRs are lower) than others. Furthermore, the areas under the ROC curves using Figure 4 were calculated as depicted in Figure 5 and Supplementary Information Figure S6. The inlet tables represent the AUC values used to draw the plots. The AUC decreases as the network density increases, and it becomes less than 60% when the network density becomes 15%, except for BY01 whose AUC already becomes 60% when the network density is 5%. Consequently, Storey01 and STS04 have the highest AUC, while BY01 has the lowest AUC, among the six FDR methods.

3.2. Network construction using real biological data

As for real experimental data, we only considered the number of discovery since the true edges among metabolites extracted from liver are unknown. Figure 6 shows the number of discovery according to different FDR methods and their cut-off values. Not surprisingly, the general trend is very similar to the simulation studies (Supplementary Information Figure S4). In detail, STS04 has the largest discovery number, followed closely by BH00. A little behind them are Story01 and BH95, which are almost perfectly overlapped with each other. BY01 is the most conservative FDR method out of the six, discovering the least number of edges.

Table 3 shows the comparison among six methods at the three cut-off values of 0.1, 0.2, and 0.3 in terms of the number of significant edges. The numbers in the upper triangle of the table represent the common discoveries present between two methods. One interesting fact is that the small set of significant edges (discoveries) is always a subset of the large set. For example, the significant edges of BY01 are always present in the set of significant edges of other methods.

To further evaluate the performance of each FDR method, we investigated whether each significant edge exists in the KEGG pathway database. The results are summarized in Table 4. In general, the number of edges present in KEGG increases as the number of significant edges increases, resulting that BH95, Storey01, BH00, and STS04 have the large number of significant edges present in KEGG like what we have seen in Figure 6 and Table 3. On the

other hand, BY01 performs the best in terms of PPV across all the cut-off values under the assumption that all interactions in KEGG pathway database are considered as true association in the real biological data. That is, the PPVs of BY01 are 73.68 % (= 14/19), 78.26 % (= 18/23), and 76.92 % (= 20/26), while these of BH95, Storey01, BH00, and STS04 are less than 60%. This is consistent with the results of simulated data when the network density is less than or equal to 5%, as can be seen in Figure 2. The constructed networks of the real metabolite data with the cut-off values of 0.1 and 0.3 can be found in Supplementary Information Figures S7 and S8, and other results, such as the list of compounds, the list of KEGG pathways, and the numbers of KEGG pathways present in the networks, are displayed in Supplementary Information Tables S2-S4.

4. Discussion and Conclusions

We evaluated the effect of six FDR methods on detecting the significant edges when constructing metabolite association network. In addition, we also examined the influence of the network structure on the performance of FDR control. To do this, we used a simulated dataset and a real metabolomics dataset. From the simulated data, we can see that BH95, BH00, Storey01, and STS04 show similar trends in terms of all eight evaluation criterion, while LFDR and BY01 hold similar behavior. In particular, in terms of F1 score and AUC, both STS04 and Storey01 perform better than the other four methods. However, BY01 performs the best among the six FDR methods in terms of PPV at the density of 5% or less. Interestingly, these findings from the simulated data can still be seen in the real metabolomics experimental data. Namely, STS04 discovers the most number of significant edges, while BY01 has the largest PPV based on the number of significant edges present in KEGG pathways database.

It is known that FDR control can be influenced by the relationship among variables, i.e., by the dependency of the p-values. Various FDR methods were developed to take care of this dependency properly. In this regard, we were interested in the effect of the network structure on FDR control for network construction since the network structure is directly related to the dependency of variables. In fact, the simulated data show a little but not significant difference between the random network and the scale-free network. That is the trend of the random network with density of 5% resembles that of the scale-free network with density of 2.5% (Figures 1-3). This might be because the scale-free network has a few hub variables with much more edges than other variables and the complexity of these hub variables are much larger than the overall complexity of the network. As a result, the practical dependency of the p-values of the scale-free network is a little higher than what the complexity is supposed to have. Nevertheless, we can see no significant effect of the network structure on the performance of FDR control, indicating that controlling FDR is likely to be independent of the topological property of connection among variables.

Both simulated and real metabolite data sets show an apparent distinction between (BH95, BH00, Storey01, STS04) and (LFDR, BY01) in their performance. This distinction gives us an obvious guideline to network construction. That is, if one is more interested in a higher PPV, the choice should be either BY01 or LFDR. One of (BH95, BH00, Storey01, STS04) should be chosen for network construction if one is concerned with either TPR or the overall

performance, such as F1 scores. Another interesting point is that the trend of BH00 in this study is a little different from that of [8], in which BH00 is the most optimal under dependence. This suggests that the p-values of GGM may have different properties with these of biomarker discovery.

The estimation of either π_0 or m_0 is a key component for BH00, Storey01, and STS04. In fact, if the value π_0 is overestimated, BH00, Storey01, and STS04 are highly likely to overestimate FDR values. We, therefore, examined the quality of the estimates of the value π_0. As the network complexity increases, the quality of the estimates becomes worse, as can be seen in Figure 7 for random networks and Supplementary Information Figure S9 for scale-free networks. The acceptable estimates can be obtained only when the network density is less than or equal to 5%. Besides, the trend of BH00 is far from that of Storey01 and STS04, and the estimate of BH00 is the worst among these three approaches, explaining why BH00 performs different from others.

In conclusion, the results suggest that STS04 should be considered to achieve the highest TPR or F1 score, while BY01 would be a choice to obtain the highest PPV in network construction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NSF grant DMS-1312603, NIH grant 1RO1GM087735, and NIH grant R21ES021311. The Biostatistics Core is supported, in part, by NIH Center Grant P30 CA022453 to the Karmanos Cancer Institute at Wayne State University.

References

1. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000; 28(1):27–30. [PubMed: 10592173]
2. Zhou H, Jin J, Zhang H, Yi B, Wozniak M, Wong L. IntPath--an integrated pathway gene relationship database for model organisms and important pathogens. BMC Systems Biology. 2012; 6:s2.
3. Whittaker, J. Wiley series in probability and mathematical statistics Probability and mathematical statistics. Wiley; Chichester England ; New York: 1990. Graphical models in applied multivariate statistics; p. 448xiv
4. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological. 1995; 57(1):289–300.
5. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98(9):5116–5121. [PubMed: 11309499]
6. Kramer N, Schafer J, Boulesteix AL. Regularized estimation of large-scale gene association networks using graphical Gaussian models. Bmc Bioinformatics. 2009; 10
7. Pihur V, Datta S, Datta S. Reconstruction of genetic association networks from microarray data: a partial least squares approach. Bioinformatics. 2008; 24(4):561–568. [PubMed: 18204062]
8. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008; 9(3):432–441. [PubMed: 18079126]

9. Benjamini Y, Hochberg Y. On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. Journal of Educational and Behavioral Statistics. 2000; 25(1):60–83.
10. Efron B, et al. Empirical Bayes analysis of a microarray experiment. Journal of the American Statistical Association. 2001; 96(456):1151–1160.
11. Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. Annals of Statistics. 2003; 31(6):2013–2035.
12. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics. 2001; 29(4):1165–1188.
13. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2004; 66:187–205.
14. Opgen-Rhein R, Strimmer K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. Bmc Systems Biology. 2007; 1
15. Tenenhaus A, et al. Gene Association Networks from Microarray Data Using a Regularized Estimation of Partial Correlation Based on PLS Regression. Ieee-Acm Transactions on Computational Biology and Bioinformatics. 2010; 7(2):251–262.
16. Jeong H, et al. The large-scale organization of metabolic networks. Nature. 2000; 407(6804):651–654. [PubMed: 11034217]
17. Efron B. Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. Journal of the American Statistical Association. 2004; 99(465):96–104.
18. Kim KI, de Wiel MAV. Effects of dependence in high-dimensional multiple testing problems. Bmc Bioinformatics. 2008; 9
19. Erdös P, Rényi A. On random graphs, I. Publicationes Mathematicae (Debrecen). 1959; 6:290–297.
20. Barabasi AL, Albert R. Emergence of scaling in random networks. Science. 1999; 286(5439):509–512. [PubMed: 10521342]
21. Albert R. Scale-free networks in cell biology. Journal of Cell Science. 2005; 118(21):4947–4957. [PubMed: 16254242]
22. Newman MEJ. Scientific collaboration networks. I. Network construction and fundamental results. Physical Review E. 2001; 64(1)
23. Adamic LA, et al. Power-Law distribution of the World Wide Web. Science. 2000; 287(5461)
24. Takahashi DY, et al. Discriminating Different Classes of Biological Networks by Analyzing the Graphs Spectra Distribution. Plos One. 2012; 7(12)
25. Dudoit S, Shaffer JP, Boldrick JC. Multiple hypothesis testing in microarray experiments. Statistical Science. 2003; 18(1):71–103.
26. Storey JD. A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2002; 64:479–498.
27. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(16):9440–9445. [PubMed: 12883005]
28. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics. 1979; 6(2):65–70.
29. Hochberg Y. A Sharper Bonferroni Procedure for Multiple Tests of Significance. Biometrika. 1988; 75(4):800–802.
30. Hommel G. A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test. Biometrika. 1988; 75(2):383–386.
31. Strimmer K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics. 2008; 24(12):1461–1462. [PubMed: 18441000]
32. Benjamini Y, Liu W. A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. Journal of Statistical Planning and Inference. 1999; 82(1-2): 163–170.

33. Blanchard G, Roquain E. Two simple sufficient conditions for FDR control. Electronic Journal of Statistics. 2008; 2:963–992.

34. Wei XL, et al. MetSign: A Computational Platform for High-Resolution Mass Spectrometry-Based Metabolomics. Analytical Chemistry. 2011; 83(20):7668–7675. [PubMed: 21932828]

35. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993; 39(4):561–77. [PubMed: 8472349]

36. Faraggi D, Reiser B. Estimation of the area under the ROC curve. Stat Med. 2002; 21(20):3093–106. [PubMed: 12369084]

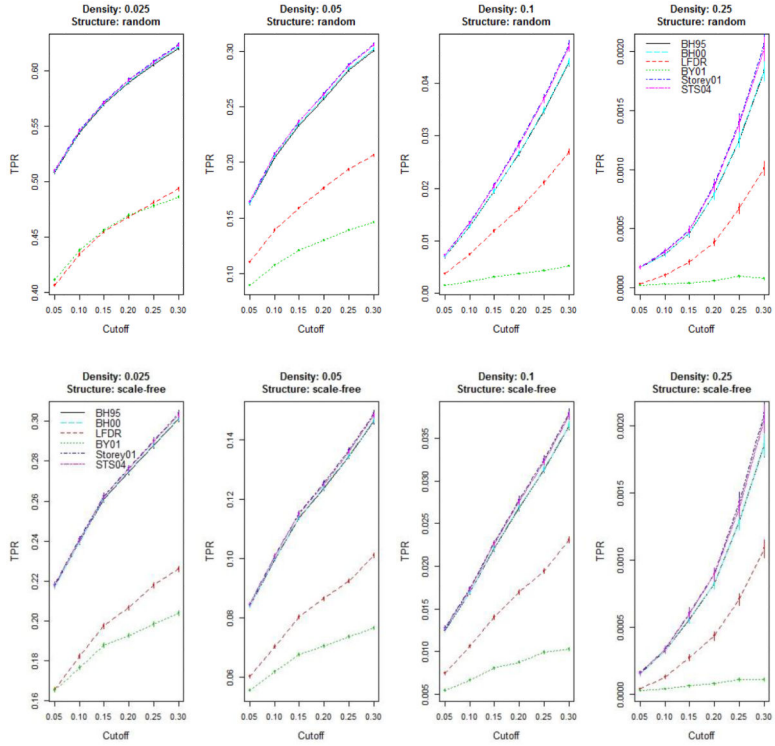


Fig 1.

The true positive rate (TPR) plots of the simulated data. The upper row represents the plots of the random networks and the plots of the scale-free networks are in the bottom row. Note that the solid bars in each line represent the standard error.

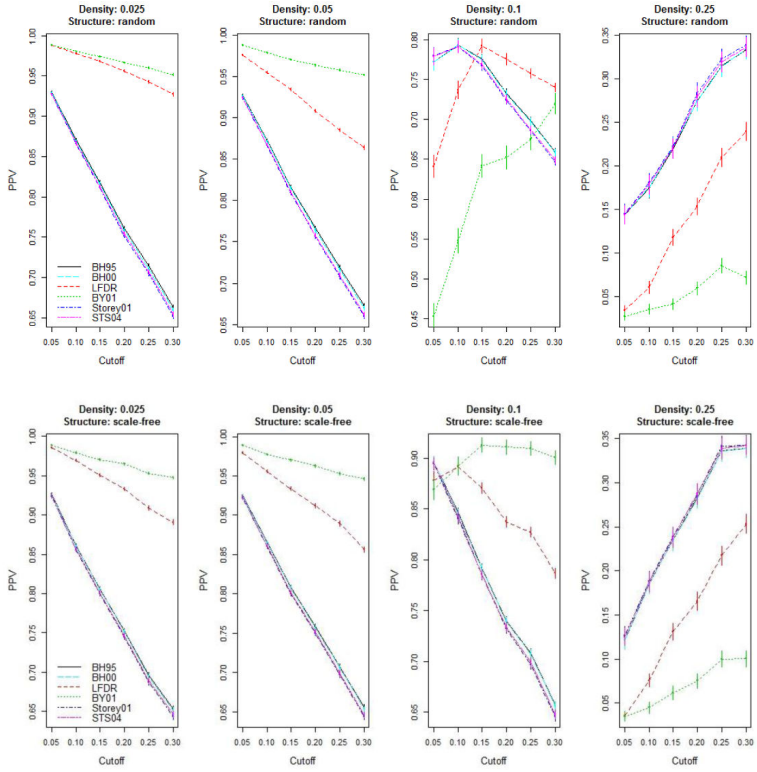


Fig 2.

The positive predictive value (PPV) plots of the simulated data. The upper row represents the plots of the random networks and the plots of the scale-free networks are in the bottom row. Note that the solid bars in each line represent the standard error.

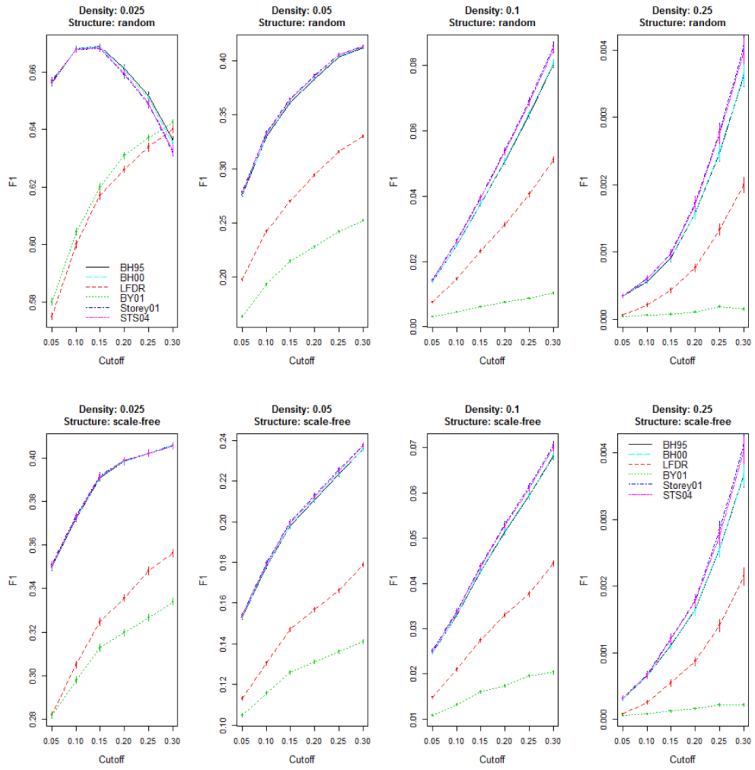


Fig 3.

The F1 score plots of the simulated data. The upper row represents the plots of the random networks and the plots of the scale-free networks are in the bottom row. Note that the solid bars in each line represent the standard error.

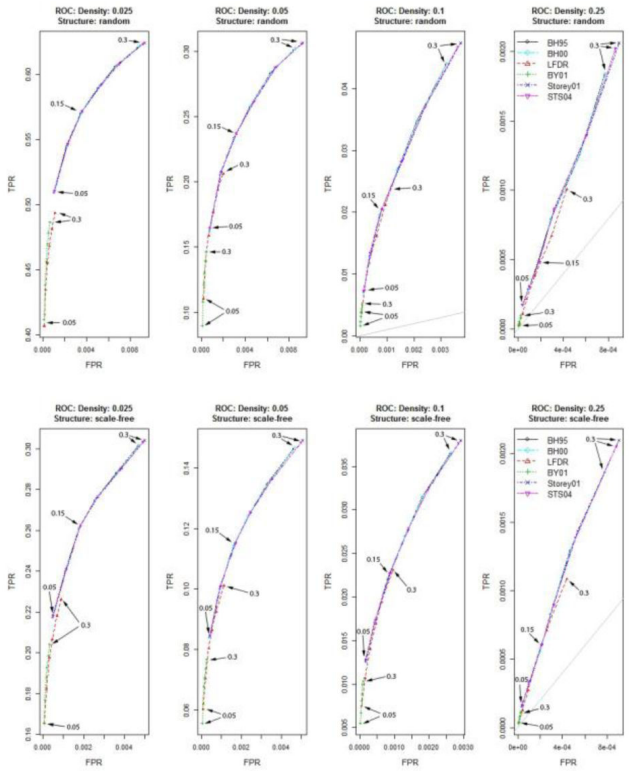
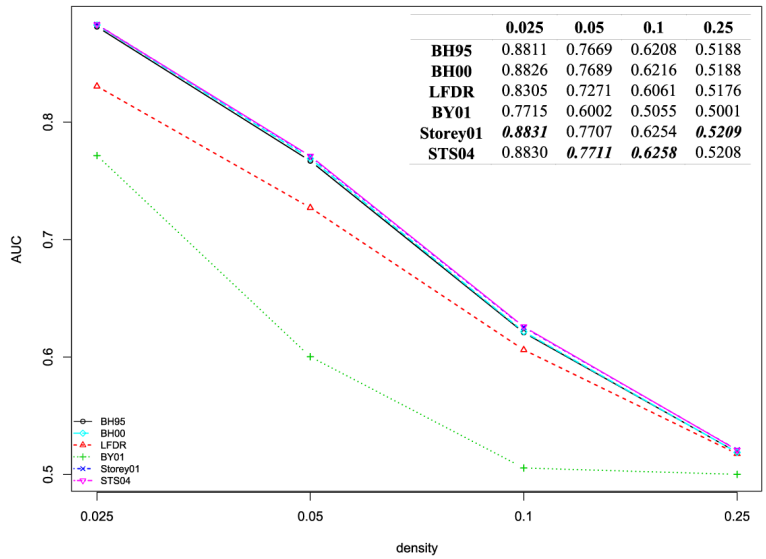


Fig 4.

The ROC plots of the simulated data. The upper row represents the plots of the random networks and the plots of the scale-free networks are in the bottom row.



	0.025	0.05	0.1	0.25
BH95	0.8811	0.7669	0.6208	0.5188
BH00	0.8826	0.7689	0.6216	0.5188
LFDR	0.8305	0.7271	0.6061	0.5176
BY01	0.7715	0.6002	0.5055	0.5001
Storey01	***0.8831***	0.7707	0.6254	***0.5209***
STS04	0.8830	***0.7711***	***0.6258***	0.5208

Fig 5.

The area under the ROC curve (AUC) plots of the simulated data for the random network. The inlet table represents the AUC values used to draw the plot.

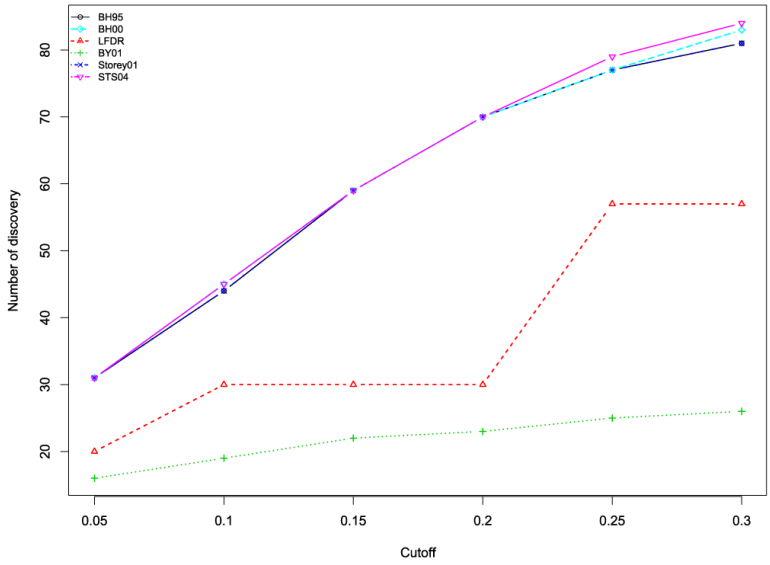


Fig 6.
The number of discovery for real experimental data.

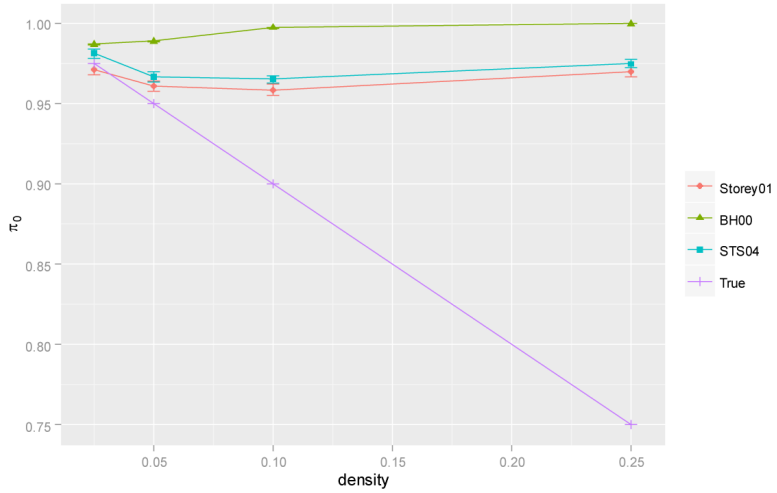


Fig 7.

Estimated $\hat{\pi}_0$ of Storey01, BH00, and STS04 versus the network density for the random network.

Table 1

Contingency table of hypothesis test

	Declared non-significant	Declared significant	Total
Null hypothesis (H_0)	U	V	m_0
Alternative hypothesis (H_1)	T	S	$m - m_0$
Total	$m - R$	R	m

Table 3

Number of intersection of significant edges by six FDR methods using real experimental data. The numbers in bold and italic indicate the maximum of each cut-off value.

Cut-off	FDR	BH95	BH00	LFDR	BY01	Storey01	STS04
0.1	**BH95**	44	44	30	19	44	44
	BH00		***45***	30	19	44	45
	LFDR			30	19	30	30
	BYO1				19	19	19
	Storey01					44	44
	STS04						***45***
0.2	**BH95**	***70***	70	30	23	70	70
	BH00		***70***	30	23	70	70
	LFDR			30	23	30	30
	BY01				23	23	23
	Storey01					***70***	70
	STS04						***70***
0.3	**BH95**	81	81	57	26	81	81
	BH00		83	57	26	81	83
	LFDR			57	26	57	57
	BY01				26	26	26
	Storey01					81	81
	STS04						***84***

Table 4

The number of KEGG pathways present in reconstructed networks for real experimental data. The numbers in bold and italic indicate the maximum number of edges present in KEGG pathways (the empirical number of discovery), and the underlined numbers represent the maximum ratio (the empirical predictive positive value (PPV)) for each cut-off value.

Cut-off	BH95	BH00	LFDR	BY01	Storey01	STS04
0.1	26/44	***27/45***	22/30	14/19	26/44	***27/45***
0.2	***39/70***	***39/70***	22/30	18/23	***39/70***	***39/70***
0.3	42/81	43/83	32/57	20/26	42/81	***44/84***