

Multiple Testing in Large-Scale Contingency Tables: Inferring Pair-Wise Amino Acid Patterns in β -Sheets

Seoung Bum Kim and Kwok-Leung Tsui

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA
E-mail: sbkim@isye.gatech.edu E-mail: ktsui@isye.gatech.edu

Mark Borodovsky

Schools of Biology and Biomedical Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA
E-mail: mb56@mail.gatech.edu

July 12, 2005

ABSTRACT

One of the most common test procedures using two-way contingency tables is a test of independence between two categorizations. Current significant tests such as χ^2 tests or likelihood ratio tests provide overall independency but bring limited information about the nature of the association in the contingency tables. This study examines the feasibility of using multiple testing procedures for an inference of independence of categories in each cell in contingency tables. In the simulation study, we compare the performance of various multiple testing procedures in a contingency table setup and demonstrate the relationship among the proportion of true null hypothesis, type I error, power, and false discovery rate. Finally, we apply the proposed methodology to identify the patterns of pair-wise associations of amino acids involved in β -sheet bridges in proteins. We identify a number of amino acid pairs that exhibit either strong or weak association. These patterns provide useful information for algorithms for predicting the secondary as well as tertiary structure of proteins.

Keywords: β -strands; Contingency table; False discovery rate; Multiple testing.

1. Introduction

One of the most common test procedures applied to two-way contingency tables is a test of independence (or association) between two categorizations. In general, the test of independence uses χ^2 tests or likelihood ratio tests that can be called “globally significant tests.” The basic idea of these tests is as follows: If the sum of all the differences between observed and expected frequencies of all cells in a contingency table is small in a statistical sense, independence between two categorizations is accepted; if the sum of the differences is large, independence is rejected. However, the global tests can hardly identify the independence of individual cells in a contingency table since their statistics are constructed based on all cells. The issue of identification of independence in individual cells is especially important in large-scale contingency tables where the number of cells $\gg 4$. Agresti (2002) pointed out several limitations of the global tests. He reviewed follow-up methods to global tests such as a partitioning of the χ^2 method as well as a method based on standardized and adjusted residual that allows further investigation of the associations in contingency tables. Partitioning of χ^2 is a method for exploring the associations by dividing the large tables into smaller ones. Lancaster (1949) showed that any $r \times c$ table can be reduced to $(r-1) \times (c-1)$ independent 2×2 tables. Hence, the interpretation of small tables is straightforward. In large-scale contingency tables, however, this method becomes too complicated as it generates too many 2×2 tables. For example, a 10×10 table produces 81 tables of 2×2 size, which makes the extraction of meaningful information cumbersome. Haberman (1973) defined the Standardized and Adjusted Residual (STAR) statistic for each cell and showed that this statistic is asymptotically standard normal under the null hypothesis of independence of category in individual cells. Therefore, the STAR statistics that are greater or less than a certain threshold indicate lack of fit to the null distribution in that cell (Agresti, 2002). The STAR method is simple but does not provide an objective way to determine a threshold since the threshold depends upon the number of degrees of freedom in a contingency table. Also, under the simultaneous consideration of all cell in contingency tables, the STAR method produces many false positives (Agresti, 2002). Another method

was also introduced by Haberman (1973), who utilized a normal probability plot of STAR values that provides a nice graphical representation. However, the interpretation of a normal probability plot is frequently subjective, particularly when the number of cells to be tested is large. Therefore, there is a need for a method able to systematically and objectively identify the independence of each cell in contingency tables. In this study, we propose a procedure for testing independence of categories in individual cells of a contingency table based on a multiple testing framework.

In multiple testing problems, family-wise error rates have been used under simultaneous consideration to avoid the multiplicity effect. Applying the single testing procedure to the multiple testing problem leads to an exponential increase of false positive rates. More precisely, the probability that at least one of the tests leads to rejection of H_0 when H_0 holds increases exponentially with the number of hypotheses. A convenient new definition of error rate, called False Discovery Rate (FDR) was proposed by Benjamini and Hochberg (1995). The FDR is the expected proportion of false positives among all the hypotheses rejected. The FDR has been used for microarray analysis to find co-expressed genes (Tusher *et al.* (2001), Efron *et al.* (2001), Efron and Tibshirani (2002), Dudoit *et al.* (2003)) as well as the genetic study to identify drugs causing mutations in the viral genome (Efron, 2004). As an extension of original FDR, Storey (2002, 2003) and Storey *et al.* (2004) introduced the positive False Discovery Rate (pFDR) and Efron *et al.* (2001) proposed the Local False Discovery Rate (Local FDR). Moreover, the case when the hypotheses are dependent was considered by Yekutieli and Benjamini (1999) and Benjamini and Yekutieli (2001).

We first review some of the multiple testing procedures and presents its application to the statistical inference of individual cells in contingency tables, the main topic of this paper. In addition, we perform simulation studies to compare the proportion of true null hypothesis, type I error, power, and FDR of different multiple testing procedures in contingency tables. Finally, the proposed procedure is applied to identify the patterns of pair-wise associations of amino acids involved in β -sheet bridges.

2. Control Procedures in Multiple Testing

2.1 The Family-Wise Error Rate

In a multiple hypothesis test, assessing the number of false positives is necessary because a mere use of single inference procedures results in a significant number of false positives (Benjamini and Hochberg, 1995). Table 1 shows the possible outcomes from m hypothesis tests. The Family-Wise Error Rate (FWER), which has been classically used as a compound error rate in the setup of multiple hypothesis testing, is defined as the probability of generating one or more false rejections, i.e.,

$$\text{FWER} = \Pr[V \geq 1], \quad (1)$$

where V is the number of rejected hypotheses when the hypothesis is true. Shaffer (1995) summarized a variety of methods controlling the FWER. The most widely used one is the Bonferroni method. This method rejects H_i if $p_i \leq \alpha_i$, where p_i is the p -value of the i th hypothesis (i.e., H_i). In general, α_i is determined equally for all hypotheses (e.g., $\alpha_i = \frac{\alpha}{m}$). Therefore, the overall FWER is less than or equal to α . Other family-wise methods were developed to improve the power of the Bonferroni method, but they are still too stringent to detect false hypotheses. In other words, they can hardly reject the null hypothesis when it is actually false. In particular, the power significantly decreases as the number of hypotheses increases, where the power is the proportion of false null hypotheses which are correctly rejected.

2.2 The False Discovery Rate

Benjamini and Hochberg (1995) introduced the False Discovery Rate (FDR), defined as the expected proportion of false positives out of all rejected null hypotheses. The advantage of the FDR is to identify as many significant hypotheses as possible while keeping a relatively small number of false positives (Storey and Tibshirani, 2003). With a large family of hypotheses, the advantages over the FWER are substantial. In Table 1, R is the number of rejected null hypotheses, and V is the

number of falsely rejected null hypotheses. Then the FDR is defined as

$$E \left[\frac{V}{V+S} \right] = E \left[\frac{V}{R} \right]. \quad (2)$$

Several important properties of the FDR were discussed in Benjamini and Hochberg (1995). For instance, R should be positive; if it is not, $\frac{V}{R}$ cannot be defined. A more exact definition of the FDR is

$$E \left[\frac{V}{R} | R > 0 \right] P(R > 0). \quad (3)$$

Understanding the relationship between the FDR and the FWER is important. When $m_0=m$, the FDR is equivalent to the FWER. When $m_0 < m$, FDR has more power in the sense that the FDR is less stringent in the multiple testing procedure. Benjamini and Hochberg (1995) proved that an ordered p -value method controls the specified FDR. This method is implemented as follows:

Consider a series of null hypotheses that are tested simultaneously,

$$H_1, H_2, \dots, H_m.$$

We denote the corresponding independent test statistics, p -values, and ordered p -values as

$$Y_1, Y_2, \dots, Y_m,$$

$$P_1, P_2, \dots, P_m,$$

$$P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}.$$

1. For a fixed α , where $0 \leq \alpha \leq 1$.
2. $\hat{i} = \max \left[i : P_{(i)} \leq \frac{i}{m} \cdot \frac{\alpha}{\pi_0} \right]$.
3. If $\hat{i} \geq 1$, $\Omega \in \{ \text{All rejected } H_i \text{ with } P_i \leq P_{(\hat{i})} \}$ with $\text{FDR}(\Omega) \leq \alpha$.
If $\hat{i} = 0$, Do not reject any hypothesis since $\Omega = \emptyset$.

Let $\pi_0 (= \frac{m_0}{m})$ denote the proportion of true H_i . In general, $\pi_0 = 1$ is the most conservative possible choice. Several studies discussed the estimate of π_0 (Storey and Tibshirani, 2003; Efron, 2004).

2.3 The Positive False Discovery Rate

Storey (2002, 2003) introduced the positive False Discovery Rate (pFDR). The term “*positive*” is included because it assumes that at least one significant hypothesis would occur.

$$\text{pFDR} = \text{E} \left[\frac{V}{R} \mid R > 0 \right]. \quad (4)$$

In terms of the controlling procedure, the Storey procedure is different from the procedure of the Benjamini-Hochberg. The latter fixes α first and then derives a rejection rule (or decision rule) that achieves $\text{FDR} \leq \alpha$ while the former fixes the rejection rule first and then estimates FDR based on this rejection rule. A detailed description of Storey’s procedure including estimation of the pFDR can be found in Storey (2002). Here, we describe the Storey’s controlling procedure briefly.

1. Reject all H_i with $P_i \leq P_{(\hat{i})}$ such that $\hat{i} = \max \left[i : \text{p}\hat{\text{FDR}}_\lambda(P_{(i)}) \leq \alpha \right]$.
2. Estimate $\text{p}\hat{\text{FDR}}_\lambda(P_{(i)})$, which is less than α .
3. Thus, $\hat{i} = \max \left[i : \text{p}\hat{\text{FDR}}_\lambda(P_{(i)}) \leq \alpha \right] = \max \left[i : \frac{\hat{\pi}_0(\lambda)mP_{(i)}}{i} \leq \alpha \right] = \max \left[i : P_{(i)} \leq \frac{i}{m} \cdot \frac{\alpha}{\hat{\pi}_0(\lambda)} \right]$.

λ , a part of the estimate of π_0 (or $\hat{\pi}_0$), is determined via a tradeoff between bias and variance (Storey, 2003). Note that the Storey’s procedure is the same as that of the Benjamini-Hochberg’s, except for its estimate of π_0 . The relationship of the two procedures was described in Storey (2002) who has shown that the two procedures are equivalent when $\hat{\pi}_0 = 1$. However, if $\hat{\pi}_0 < 1$ and $\hat{\pi}_0$ can be properly estimated, the Storey’s procedure provides more power while controlling the same FDR. In other words, if the Storey’s and Benjamini-Hochberg’s procedures reject the same number of hypotheses, the Storey’s procedure has a smaller FDR (Storey, 2002).

2.4 The Local False Discovery Rate

Efron *et al.* (2001) introduced the Local False Discovery Rate (Local FDR), the empirical Bayes version of the original FDR. Suppose the test statistics from multiple hypotheses follow a mixture distribution of two classes, i.e., statistics for true null and false null. Prior probabilities and their

corresponding densities are represented as follows:

$$\pi_0 = \text{probability of true null}, \quad f_0(y) = \text{the density of } Y \text{ for true null.}$$

$$\pi_1 = \text{probability of false null}, \quad f_1(y) = \text{the density of } Y \text{ for false null.}$$

Then the mixture density can be expressed as

$$f(y) = \pi_0 f_0(y) + \pi_1 f_1(y). \quad (5)$$

Given y , the posterior probabilities of being in either the true null class or the false null class are as follows:

$$\Pr\{\text{true null}|y\} = \frac{\pi_0 f_0(y)}{f(y)}, \quad (6)$$

$$\Pr\{\text{false null}|y\} = 1 - \Pr\{\text{true null}|y\} = \frac{\pi_1 f_1(y)}{f(y)}. \quad (7)$$

The Local FDR is defined to be

$$\text{Local FDR}(y) = \frac{\pi_0 f_0(y)}{f(y)}, \quad (8)$$

where $\pi_0 = 1$ gives the upper bound of the Local FDR. Efron (2004) suggested the following procedure to identify significant hypotheses.

1. Estimate $f(y)$ from test statistics, say $\hat{f}(y)$.
2. Estimate a null density $f_0(y)$, say $\hat{f}_0(y)$.
3. Estimate π_0 , say $\hat{\pi}_0$.
4. Compute the Local FDR(y) = $\frac{\hat{\pi}_0 \hat{f}_0(y)}{\hat{f}(y)}$.
5. Declare y significant if Local FDR(y) $\leq \delta$, where δ is some threshold value.

The Local FDR, as its name suggests, provides a measure for the *specific (or local)* hypothesis by taking the ratio of true null density to mixture density for each set of test statistics. Thus, the small value of the ratio (i.e., $f(y)$ is much larger than $\pi_0 f_0(y)$) implies a high chance that the

hypothesis with statistic y is false. Efron and Tibshirani (2002) showed that the Local FDR has a close relationship with Benjamini and Hochberg's FDR. The conditional expectation of the Local FDR given a rejection region is the same as the Benjamini and Hochberg's FDR.

3. Multiple Testing in Contingency Tables

Our main interest is statistical inference of independence of categories in each cell in contingency tables. Table 2 presents a two-way contingency table.

Usually, χ^2 tests or likelihood ratio tests have been used to identify the association of two categorizations under the null hypothesis of independence, i.e.,

$$H_o : p_{ij} = p_{i*} \cdot p_{*j}, \quad (9)$$

$$i = 1, 2, \dots, r, \quad j = 1, 2, \dots, c.$$

Pearson's χ^2 and the likelihood ratio test statistics, i.e., L^2 , are defined in Equation 10 where N_{ij} and E_{ij} are the observed and expected values in a cell corresponding to the i th row and the j th column.

$$X^2 = \sum_i \sum_j \frac{(N_{ij} - E_{ij})^2}{E_{ij}}, \quad L^2 = 2 \sum_i \sum_j N_{ij} \log \left(\frac{N_{ij}}{E_{ij}} \right). \quad (10)$$

We call the χ^2 and the likelihood ratio tests "*globally significant tests*", as these test statistics are derived from the sum of the deviations in all cells, i.e., $\sum_i \sum_j (N_{ij} - E_{ij})$. These global tests can evaluate overall association for the two categorizations in contingency tables but give little information about individual cells. Cochran (1954) and Berkson (1938) warned that the unguarded use of globally significant tests can mislead decision makers. For instance, if the χ^2 tests accept the null hypothesis, one should conclude that no significant association exists between two categorizations. However, some cells can have large deviations between N_{ij} and E_{ij} and the χ^2 tests fail to identify those cells that may contain useful information. Therefore, to find each important cell in contingency tables, appropriate methods need to be developed.

We consider now contingency tables as a data set to test multiple hypotheses simultaneously. More precisely, for an $r \times c$ contingency table, we have following $r \times c$ hypotheses.

$$\begin{aligned} H_1 : p_{11} &= p_{1*} \cdot p_{*1} \\ H_2 : p_{12} &= p_{1*} \cdot p_{*2} \\ &\dots \\ H_{r \times c} : p_{rc} &= p_{r*} \cdot p_{*c}. \end{aligned}$$

Under the null hypothesis of independence, the adjusted residual (e_{ij}) for each cell can be defined as follows.

$$e_{ij} = \frac{N_{ij} - \frac{N_{i*} \cdot N_{*j}}{N_{**}}}{\left(\frac{N_{i*} \cdot N_{*j}}{N_{**}}\right)^{\frac{1}{2}}} = \frac{N_{ij} - E_{ij}}{(E_{ij})^{\frac{1}{2}}}. \quad (11)$$

Haberman (1973) proved that

$$e_{ij} \xrightarrow{\mathcal{D}} \mathcal{N}(0, v_{ij}), \quad (12)$$

where

$$v_{ij} = \left(1 - \frac{N_{i*}}{N_{**}}\right) \left(1 - \frac{N_{*j}}{N_{**}}\right). \quad (13)$$

However, asymptotic variance of e_{ij} is less than or equal to 1 unless the sample size is large enough (Haberman, 1973). A Standardized and Adjusted Residual (STAR), derived from dividing e_{ij} by its standard error, has been utilized as a corrected statistic. Under H_0 , STAR values follow an asymptotic standard normal distribution.

$$\tilde{e}_{ij} = \frac{N_{ij} - E_{ij}}{(E_{ij} v_{ij})^{\frac{1}{2}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \quad (14)$$

The complete derivation can be found in Haberman (1973) and Agresti (2002). We used \tilde{e}_{ij} as a test statistics for each cell in a contingency table. Agresti (2002) mentioned that the absolute value of \tilde{e}_{ij} , which exceeds about 2 (or in some cases 3), indicates the significant difference between observed and expected frequencies in that cell. Additionally, Haberman (1973) suggested

the normal probability plotting of \tilde{e}_{ij} values for identifying lack of independence of cells. Thus, the \tilde{e}_{ij} that significantly deviates from the straight line is interpreted as an indicator of strong association between categories i and j . However, as we mentioned earlier, the methods described above are rather subjective thus do not provide an objective measure of large deviation. Below we propose a multiple testing procedures for contingency tables to identify the individual cells that are significantly associated between categories. The proposed procedure is summarized as follows:

Summary of the proposed procedure (multiple testing in a contingency table)

Consider a two-way $r \times c$ contingency table,

1. Construct $r \times c$ hypotheses for testing the independence of each cell.

$$H_{ij} : p_{ij} = p_{i*} \cdot p_{*j} \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, c.$$

2. Compute the corresponding test statistics based on STAR.

$$\tilde{e}_{ij} = \frac{N_{ij} - E_{ij}}{(E_{ij}v_{ij})^{\frac{1}{2}}} \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, c.$$

Under the null hypothesis, $\tilde{e}_{ij} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$.

3. Choose one of the multiple testing procedures described in Section 2 to identify the significant cells in a contingency table.

4. Simulation Studies

4.1 The Setting

We apply our proposed procedure to the simulated contingency tables. We compare the empirical power, type I error, and false discovery rate of four different multiple testing procedures and the individual test under different scenarios. Details are as follows:

1. Bonferroni procedure controlling FWER at 0.01.
2. Benjamini and Hochberg procedure controlling FDR at 0.01.
3. Storey procedure controlling at pFDR at 0.01.
4. Efron procedure with the threshold of local FDR at 0.01.
5. Individual test with the p -value threshold at 0.01.

To clarify, we define the empirical power, type I error, and false positive rate as follows:

$$\begin{aligned} \text{Empirical power} &= \frac{\# \text{ of correctly rejected hypotheses}}{\# \text{ of false null hypotheses}}, \\ \text{Empirical type I error} &= \frac{\# \text{ of incorrectly rejected hypotheses}}{\# \text{ of true null hypotheses}}, \\ \text{Empirical FDR} &= \frac{\# \text{ of incorrectly rejected hypotheses}}{\# \text{ of rejected hypotheses}}. \end{aligned}$$

In this study we consider a 4×4 contingency table. That is, 16 hypotheses (=total number of cells) are considered for testing independence of categories in each cell. Thus, the family of null hypotheses states that 4×4 categories in the contingency table are independent.

$$H_{ij} : p_{ij} = p_{i*} \cdot p_{*j} \quad i = 1, 2, 3, 4, \quad j = 1, 2, 3, 4.$$

Each individual null hypothesis is tested based on standardized and adjusted residual statistics (Equation 14), and these test statistics are assumed to be independent.

The proportion of true null hypotheses out of 16 is set to be 0%, 25%, 50%, and 75%. In other words, the number of significant ones out of a total of 16 hypotheses is 16, 12, 8, and 4, respectively. In addition, we define θ (Equation 15), which measures a magnitude of difference between p_{ij} and $p_{i*} \cdot p_{*j}$ in the contingency table. Hence, the contingency table having large θ implies the huge discrepancy between observed and expected frequencies in that contingency table. In our simulations, different θ s are considered in each scenario, where

$$\theta = \sum_{i=1}^r \sum_{j=1}^c |p_{ij} - p_{i*} \cdot p_{*j}| \quad . \quad (15)$$

Finally, we consider sample sizes of $n = 100, 500$, and 1000 to investigate their effects. Detailed procedure to generate the simulated contingency tables are not presented here due to space limit but available in email by authors.

4.2 Results

Each simulation is done with 5000 repetitions. Figures 1 ~ 4 illustrate the average empirical power, type I error, and FDR for the five different procedures (i.e., individual, Bonferroni, Benjamini-Hochberg, Storey, and Efron). In each panel of figures, the x-axis is a different value of θ and the y-axes are respectively the average empirical power, type I error, and FDR. The following observations are made.

1. Average empirical power (Figure 1)

- (a) Power of all procedures increases when both sample size and θ increase.
- (b) Generally, the individual test produces larger power as well as higher type I error and false discovery rate than the other four procedures. However, as the proportion of true null hypotheses decreases (i.e., the proportion of significant hypotheses increases), the FDR-related procedures give the power comparable to the individual test. Storey's procedure yields larger power than the individual test in some cases. For the comparison of multiple testing procedures, the power is uniformly ranked as follows:

Storey > Benjamini-Hochberg > Efron > Bonferroni.

2. Average empirical FDR (Figure 2)

- (a) FDR decreases when the proportion of true null decreases. The reason is described as follows: Using the notation of Table 1, let the number of true and false null be denoted by m_0 and m_1 . Moreover, let type I error and power be denoted by α and $(1 - \beta)$. Then the FDR can be represented as follows:

$$FDR = \frac{m_0\alpha}{m_0\alpha + m_1(1 - \beta)}. \quad (16)$$

Thus, if α and $(1 - \beta)$ are fixed, the FDR decreases as m_0 decreases. Note that as m_0 decreases, m_1 increases.

- (b) The individual test produces larger FDR than the other procedures when δ is small and the proportion of significant hypotheses increases.

3. Average empirical type I error (Figure 3)

- (a) Type I error increases as the number of true null decreases when the FDR is fixed. From Equation 16, we can derive the following equation,

$$\alpha = \frac{m_1}{m_0} \cdot \frac{FDR}{1 - FDR} \cdot (1 - \beta). \quad (17)$$

Equation 17 shows that α is inversely proportional to m_0 .

- (b) Type I error of the individual test procedure and the Bonferroni procedure are constant over each simulation scenario. Type I error of the FDR-related procedures increases as the proportion of significant hypotheses increases (i.e., $\pi_0 \rightarrow 0$). Also as the δ increases, type I error of the FDR-related procedures increases.

- 4. Figure 4 shows type I error and false discovery rate when the proportion of true null hypothesis is 100%. In this case, the power is not defined by its definition. The individual test produces larger type I error and the FDR than other procedures over the different sample sizes.

The conclusions from the simulations show that the number of significant hypotheses is small, the individual test procedure gives large power but renders high FDR. However, if the number of significant hypotheses is large, the Storey's procedure produces large power with relatively small type I error compared to the individual test procedure.

The simulation results here with contingency tables agree well with the previous simulation studies assuming a normal random variable (Storey, 2002). This is due to the asymptotic normal assumption of individual cells in contingency tables. Other studies mainly focus on the comparison of power for different multiple testing procedures but we conduct more comprehensive studies.

5. Identification of Significant Amino Acid Pairs in Protein β -Sheets

5.1 Background

In this section we apply to our proposed procedure for the identification of the patterns of pair-wise association involved in β -sheet bridges defined in Kabsch and Sander (1983). Pair-wise association arises when two amino acids are distantly located in a primary structure but they are close in the tertiary structure because of protein folding. Knowledge of these patterns can improve the prediction accuracy of protein structures. For instance, the secondary protein prediction problem involves predicting the location of α -helices, β -sheets, and loops from a primary structure of proteins. It has been suggested that existing methods achieve reasonably accurate prediction rate for identifying α -helices and loops but prediction rate of β -sheets remains significantly low due to their pair-wise associated pattern (Frishman and Argo, 1996). Thus, investigating the pair-wise associations in β -sheets can improve overall prediction accuracy of protein secondary structure as well as provide useful information of prediction of protein tertiary structure. In the last several decades many studies have addressed this issue. Von Heijne and Blomberg (1977, 1978) studied the pair correlations in β -strands among hydrophobic, neutral, and polar classes of residues. They revealed that residues within the same classes occur more often than expected by random chance. Lifson and Sander (1980) analyzed the frequencies of amino acid pairs in parallel and antiparallel structures and uncovered the number of trends in favored amino acid pairs. But their studies were performed on the group level so the results did not provide individual patterns of pair-wise association. Recent studies focused on antiparallel β -sheets (Wouters and Curmi (1995), Smith and Regan (1995), Hutchison *et al.* (1998)). They investigated two distinct sites based on the existence of hydrogen bonding in backbone NH and C=O: hydrogen bonded and non-hydrogen bonded. Their work revealed that two sites have different patterns of residue pairs. General consensus of previous studies implies that nonrandom patterns of pair-wise associations exist across neighboring β -strands.

5.2 Problem Formulation

The primary aim of this application is to find “*avored pairs*” and “*unavored pairs*” among all possible pairs of amino acids. Here, the term “*avored*” or “*unavored*” means that two amino acids like or do not like to be associated with each other to form β -sheet bridges. The frequency of amino acid pairs in β -sheets is obtained from the 613 known proteins in Protein Data Bank (PDB). Secondary structures are assigned to the 613 proteins using the Definition of Secondary Structure Assignment algorithm (DSSP; <http://www.cmbi.kun.nl/gv/dssp/>) designed by Kabsch and Sander (1983). These data explicitly indicate which pairs of amino acids from bridges in β -sheets and allow to construct a 20×20 contingency table (Table 3). There are 20 different amino acids and thus the possible number of pairs are 400 ($= 20 \times 20$). However, we only consider 210 pairs since we do not distinguish between the two different types of amino acids in a pair. In other words, X:Y observation leads to increment by one of both N_{XY} and N_{YX} . Our main task is to identify which of the 210 pairs are significantly associated to form β -sheet bridges. In each category of the contingency table, the null hypothesis is that two amino acids are paired at random. More precisely, we can construct the following 210 hypotheses.

$$H_1 : p_{AA} = p_{A*} \cdot p_{*A},$$

$$H_2 : p_{AC} = p_{A*} \cdot p_{*C},$$

...

$$H_{210} : p_{YY} = p_{Y*} \cdot p_{*Y}.$$

5.3 Results

The results from the multiple testing procedures are reported in Tables 4 ~ 7 and summarized in Table 8. Comparing the individual test procedure with the Bonferroni’s procedure, the former with the p -value threshold of 0.01 per each hypothesis found more significant hypotheses than the latter controlling FWER=0.01. Among the procedures controlling the FDR, the Storey’s procedure found the largest number of hypotheses. The Benjamini-Hochberg and Efron procedures found the second

and the third largest. Another observation is that the Storey's procedure identifies more significant hypotheses than the individual test procedure. This is explained in the simulations. In our example, the estimated proportions of true null ($\hat{\pi}_0$ s) of the parallel and antiparallel strands are 0.2433 and 0.1818, obtained from software developed by Storey (<http://faculty.washington.edu/jstorey/>). This implies that the estimated proportion of significant hypotheses is large in our example. In this case, the Storey's procedure is desirable since the simulation study showed that it produces large power with relatively small type I error when $\hat{\pi}_0$ is small.

Another important issue between the individual test procedure and the FDR-related procedures is an interpretation. For instance, in our problem, we are more interested in the fraction of false positives among all rejected hypotheses than the probability of making one or more false positive rates.

To clarify, consider the case of significant pairs (both favored and unfavored) in antiparallel strands. The individual test found 118 significant pairs with the probability that at least one false positive is 0.88. For the Bonferroni test, 76 pairs are found to be significant with the probability that at least one false positive is 0.01. The other three methods controlling the FDR can be interpreted similarly. For example, Storey's procedure declares 130 significant pairs among which the expected proportion of false positives is 0.01. Note that Storey's procedure found more significant hypotheses (large power) than do the procedures of Benjamini and Hochberg and the Efron while controlling the same FDR.

Once we identify the significant pairs by multiple testing procedure, the next step is to interpret the results biologically. Interpretations are made based on Storey's procedure. We utilize the chemical properties of amino acids based on the following four groups (Alberts *et al.* 1997):

- Negatively charged polar: Asp(D), Glu(E).
- Positively charged polar: Arg(R), Lys(K), His(H).
- Uncharged Polar: Asn(N), Gln(Q), Ser(S), Thr(T), Tyr(Y).

- Nonpolar: Ala(A), Gly(G), Val(V), Lue(L), Ile(I), Pro(P), Phe(F), Met(M), Trp(W), Cys(C).

Parallel Strands: With regard to favored pairs (Table 4), strong associations are shown between two amino acids with opposite charges: E:K(1), D:R(2), D:K(5), E:R(12), D:H(16), E:H(23). Here, the values inside the parentheses indicate the ranking based on the STAR measure of significant association. High associated level is also observed between hydrophobic amino acids. For instance, I:I(4), I:L(6), V:V(7), M:M(17), F:V(20), and so on. Other high associations are shown within polar amino acids and between positively charged polar and uncharged polar amino acids.

Observations on pairs which exhibit a significantly unfavored pattern in parallel strands are presented in Table 5. Such pairs are rarely observed to form β -sheet bridges. Nonpolar amino acids tend not to be associated with uncharged polar amino acids. Moreover, both positively and negatively charged polar amino acids show no inclination to interact with nonpolar amino acids.

Antiparallel Strands: Tables 6 and 7 give the lists of significantly favored and unfavored pairs in antiparallel strands. Overall patterns of antiparallel strands are similar to those in parallel strands. Strong interactions between the negatively and positively charged groups are observed: E:K(1), E:R(2), D:R(5), D:H(13), D:K(21), E:H(50). Note that glutamic acid (E) and lysine (K) have the strongest interaction in both parallel and anti-parallel strands. Many hydrophobic interactions also take place: L:L(4), C:C(8), I:I(9), V:V(10), I:V(11), I:L(12), P:W(19), F:L(22), F:F(24), A:A(26), L:M(32), A:I(35), C:I(37), and so on.

For unfavored pairs, we particularly observed that nonpolar residues show the tendency not to be associated with negatively charged polar residues: D:V(205), E:I(203), D:I(198), D:L(187), E:L(183), and E:V(154).

6. Conclusion

This study presents an approach to statistical inference of independence of categories in each cell in contingency tables within the multiple testing framework. The proposed procedure compensates for the limitation of the globally significant tests such as χ^2 and likelihood ratio tests in that it

provides more information about the nature of the association in each cell in contingency tables. Moreover, the procedure has an advantage over the subjective methods such as normal probability plotting and partitioning of χ^2 . In large-scale contingency tables in particular, the proposed procedure provides an objective and systematic way of finding the significantly associated cells. In this paper, four multiple testing procedures (Bonferroni, Benjamini-Hochberg, Efron, and Storey) that control corresponding compound errors (FWER, FDR, Local FDR, pFDR) as well as individual testing testing procedure are employed and compared. The simulation studies show that the procedures controlling pFDR, FDR, and Local FDR provide higher power than BF method controlling classical FWER. The high power allows further characterization of the identified cells. For the case study, the proposed procedure has been applied to identify the patterns of pair-wise association of amino acids in β -sheet bridges and produced a list of favored and unfavored pairs. The statistical procedure considered in this paper cannot fully identify the physical or chemical nature of observed associations. However, these results are useful for better understanding of protein structure and should help develop better algorithms of the protein secondary and tertiary structure prediction.

Acknowledgements

We are grateful to Zafer Aydin for valuable help in extracting experimental data of protein secondary structure. M.B. was supported in part by NIH grant HG00783.

References

- [1] Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Robert, K., and Walter, P. (1997). *Essential Cell Biology: An Introduction to the Molecular Biology of the Cell* Garland Publishing, Inc., New York.
- [2] Agresti, A. (2002). *Categorical Data Analysis*. New York: John Wiley and Sons, Inc.
- [3] Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, **57**, 289–300.

- [4] Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.
- [5] Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *J. Am. Stat. Assoc.*, **33**, 526–536.
- [6] Cochran, W. G. (1954). Some methods of strengthening the common χ^2 tests. *Biometrics* **10**, 417–451.
- [7] Dudoit, S., Shaffer, J.P., and Boldrick, J.C. (2003). Multiple Hypothesis Testing in Microarray Experiments. *Statistical Science* **18**, 71–103.
- [8] Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Am. Stat. Assoc.*, **99**, 99–104.
- [9] Efron, B., and Tibshirani R. (2001). Microarrays, empirical Bayes methods, and false discovery rates. *Technical Report 2001-217, Department of Statistics, Stanford University*.
- [10] Efron, B., Tibshirani, R., Storey, J.D., and Tusher, V. (2001). Empirical Bayes Analysis of a Microarray Experiment. *Journal of American Statistical Association*, **96**, 1151–1160.
- [11] Frishman, D., and Argos. P. (1996). Incorporation of non-local interactions in protein secondary structure prediction. *Protein Engineering*, **9**, 133–142.
- [12] Haberman, S. J. (1973). The analysis of residuals in cross-classified tables. *Biometrics* **29**, 205–220.
- [13] Hutchinson E.G., Sessions, R.B., Thornton, and J.M., Woolfson, D. N. (1998). Determinants of strand register in antiparallel β -sheets of proteins. *Protein Science*, **7**, 2287–2300.
- [14] Kabsh, W., and Sander, C. (1983). Dictionary of proteins secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.

- [15] Lancaster, M.B. (1949). The derivation and partition of χ^2 in certain discrete distributions. *Biometrika* **36**, 117–129.
- [16] Lifson, S., and Sander, C. (1980). Specific recognition in the tertiary structure of β -sheets of proteins. *Journal of Molecular Biology* **139**, 627–639.
- [17] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, New York.
- [18] Shaffer, J. (1995). Multiple hypothesis testing. *Annu. Rev. Psychol.*, **46**, 561–584.
- [19] Smith C.K., and Regan L. (1995). Guidelines for protein design - The energetics of beta-sheet side-chain interactions. *Science*, **270**, 980–982.
- [20] Storey J.D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B*, **64**, 479–498.
- [21] Storey J.D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics* **31**, 2013–2035.
- [22] Storey J.D., Taylor, J.E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B*, **66**, 187–205.
- [23] Storey J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. NatlAcad. Sci., USA* **100**, 9440–9445.
- [24] Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarray applied to the ionizing radiation response. *Proc. NatlAcad. Sci., USA* **98**, 5116–5121.
- [25] von Heijne, G., and Blomberg, C. (1977). The β -structure: Inter-strand Correlations. *Journal of Molecular Biology*, **117**, 821–824.
- [26] von Heijne, G., and Blomberg, C. (1978). Some Global β -sheet Characteristics. *Bipolymers*, **17**, 2033–2037.

- [27] Wouters, M.A., and Curmi, P.M.G. (1995). An analysis of side-chain interactions and pair correlations within antiparallel beta-sheets: The differences between backbone hydrogen bonded and non-hydrogen-bonded residue pairs. *Proteins*, **22**, 119–131.
- [28] Yekutieli, D., and Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, **82**, 171–196.

Table 1: Outcomes from the multiple hypothesis tests of size m

	Accept null hypothesis	Reject null hypothesis	Total
True null hypothesis	U	V	m_0
False null hypothesis	T	S	m_1
Total	W	R	m

Table 2: A two-way $r \times c$ contingency table

	1	2	\dots	c	
1	N_{11}	N_{12}	\dots	N_{1c}	N_{1*}
2	N_{21}	N_{22}	\dots	N_{2c}	N_{2*}
\dots			\dots		\dots
r	N_{r1}	N_{r2}	\dots	N_{rc}	N_{r*}
	N_{*1}	N_{*2}	\dots	N_{*c}	N_{**}

Table 3: A two-way 20×20 contingency table containing the frequency of pair-wise amino acid in β -sheet bridges

	A	C	\dots	Y	
A	N_{AA}	N_{AC}	\dots	N_{AY}	N_{A*}
C	N_{CA}	N_{CC}	\dots	N_{CY}	N_{C*}
\dots			\dots		\dots
Y	N_{YA}	N_{YC}	\dots	N_{YY}	N_{Y*}
	N_{*A}	N_{*C}	\dots	N_{*Y}	N_{**}

Table 4: Favored amino acid pairs in parallel β -sheet bridges. Five procedures (Individual (IND), Bonferroni (BF), Benjamini-Hochberg (B-H), Efron (EF), and Storey (ST)) for controlling corresponding false positive rates (FPR, FWER, FDR, Local FDR, and pFDR) at $\alpha = 0.01$ are applied to find significant pairs. S and N represent significant and nonsignificant pairs

Rank	Amino acid pair	STAR	p -value	Local FDR	IND	BF	B-H	EF	ST
1	E:K	16.43	0	0	S	S	S	S	S
2	D:R	13.65	0	0	S	S	S	S	S
3	N:T	11.86	0	0	S	S	S	S	S
4	I:I	10.07	0	0	S	S	S	S	S
5	D:K	10.04	0	0	S	S	S	S	S
6	I:L	8.39	0	0	S	S	S	S	S
7	V:V	7.95	0	0	S	S	S	S	S
8	N:N	7.65	0	0	S	S	S	S	S
9	H:T	6.68	0	0	S	S	S	S	S
10	T:T	6.48	0	0	S	S	S	S	S
11	K:T	6.38	0	0	S	S	S	S	S
12	E:R	6.31	0	0	S	S	S	S	S
13	S:T	6.24	0	0	S	S	S	S	S
14	R:T	6.11	0	0	S	S	S	S	S
15	Q:W	5.59	0	0	S	S	S	S	S
16	D:H	4.94	0	0	S	S	S	S	S
17	M:M	4.85	0	0	S	S	S	S	S
18	P:T	4.75	0	0.00013	S	S	S	S	S
19	Q:Y	4.72	0	0.00014	S	S	S	S	S
20	F:V	4.62	0	0.00023	S	S	S	S	S
21	N:Q	4.53	0	0.00032	S	S	S	S	S
22	S:S	4.34	0.00002	0.00069	S	S	S	S	S
23	E:H	4.18	0.00002	0.00130	S	S	S	S	S
24	M:P	3.95	0.00008	0.00311	S	N	S	S	S
25	Y:Y	3.80	0.00014	0.00529	S	N	S	S	S
26	C:R	3.76	0.00018	0.00607	S	N	S	S	S
27	D:S	3.70	0.00022	0.00724	S	N	S	S	S
28	L:V	3.35	0.00080	0.02227	S	N	S	N	S
29	L:L	3.08	0.00204	0.04857	S	N	S	N	S
30	C:V	3.08	0.00210	0.04946	S	N	S	N	S
31	I:V	2.96	0.00308	0.06818	S	N	S	N	S
32	A:L	2.94	0.00328	0.07137	S	N	N	N	S
33	K:P	2.90	0.00374	0.08026	S	N	N	N	S
34	N:S	2.89	0.00386	0.08224	S	N	N	N	S
35	H:S	2.70	0.00688	0.13102	S	N	N	N	S
36	G:Y	2.67	0.00758	0.14167	S	N	N	N	S
37	Q:S	2.64	0.00830	0.15229	S	N	N	N	S
38	M:W	2.62	0.00876	0.15947	S	N	N	N	S
39	N:W	2.59	0.00952	0.16998	S	N	N	N	S
40	Q:T	2.56	0.01056	0.18514	N	N	N	N	S
41	H:K	2.54	0.01114	0.19299	N	N	N	N	S
42	F:G	2.46	0.01394	0.23061	N	N	N	N	S
					N	N	N	N	N
					:	:	:	:	:

Table 5: Unfavored amino acid pairs in parallel β -sheet bridges. See the caption of Table 4 for definitions of columns

Rank	Amino acid pair	STAR	p -value	Local FDR	IND	BF	B-H	EF	ST
210	I:R	-7.25	0	0	S	S	S	S	S
209	I:S	-6.94	0	0	S	S	S	S	S
208	I:N	-6.36	0	0	S	S	S	S	S
207	N:V	-6.27	0	0	S	S	S	S	S
206	L:T	-5.85	0	0	S	S	S	S	S
205	T:V	-5.70	0	0	S	S	S	S	S
204	H:I	-5.31	0	0	S	S	S	S	S
203	I:K	-5.28	0	0	S	S	S	S	S
202	R:V	-5.24	0	0	S	S	S	S	S
201	D:I	-5.19	0	0	S	S	S	S	S
200	A:T	-5.07	0	0	S	S	S	S	S
199	I:T	-4.95	0	0	S	S	S	S	S
198	E:V	-4.92	0	0	S	S	S	S	S
197	G:I	-4.77	0	0.00010	S	S	S	S	S
196	L:N	-4.76	0	0.00011	S	S	S	S	S
195	E:L	-4.75	0	0.00011	S	S	S	S	S
194	I:Q	-4.69	0	0.00015	S	S	S	S	S
193	L:S	-4.65	0	0.00017	S	S	S	S	S
192	L:Q	-4.60	0	0.00022	S	S	S	S	S
191	D:V	-4.53	0	0.00029	S	S	S	S	S
190	K:V	-4.33	0.00002	0.00066	S	S	S	S	S
189	F:K	-4.31	0.00002	0.00071	S	S	S	S	S
188	L:W	-4.26	0.00002	0.00086	S	S	S	S	S
187	H:V	-4.24	0.00002	0.00096	S	S	S	S	S
186	A:M	-4.12	0.00004	0.00151	S	S	S	S	S
185	D:F	-4.05	0.00006	0.00198	S	N	S	S	S
184	K:L	-3.99	0.00006	0.00244	S	N	S	S	S
183	F:T	-3.95	0.00008	0.00287	S	N	S	S	S
182	S:V	-3.32	0.00089	0.02354	S	N	S	N	S
181	Q:V	-3.23	0.00126	0.03167	S	N	S	N	S
180	A:N	-3.19	0.00143	0.03539	S	N	S	N	S
179	D:W	-3.10	0.00197	0.04625	S	N	S	N	S
178	E:I	-3.05	0.00227	0.05226	S	N	S	N	S
177	V:Y	-3.05	0.00231	0.05310	S	N	S	N	S
176	W:Y	-3.04	0.00237	0.05409	S	N	S	N	S
175	D:Y	-2.82	0.00476	0.09669	S	N	S	N	S
174	C:Q	-2.72	0.00646	0.12438	S	N	S	N	S
173	T:W	-2.61	0.00911	0.16468	S	N	N	N	S
172	M:T	-2.50	0.01256	0.21350	N	N	N	N	S
171	C:T	-2.41	0.01590	0.25571	N	N	N	N	S
					N	N	N	N	N
					:	:	:	:	:

Table 6: Favored amino acid pairs in antiparallel β -sheet bridges. See the caption of Table 4 for definitions of columns

Rank	Amino acid pair	STAR	p -value	Local FDR	IND	BF	B-H	EF	ST
1	E:K	21.73	0	0	S	S	S	S	S
2	E:R	18.06	0	0	S	S	S	S	S
3	N:T	13.53	0	0	S	S	S	S	S
4	L:L	12.43	0	0	S	S	S	S	S
5	D:R	12.21	0	0	S	S	S	S	S
6	T:T	12.20	0	0	S	S	S	S	S
7	S:T	10.46	0	0	S	S	S	S	S
8	C:C	9.64	0	0	S	S	S	S	S
9	I:I	9.47	0	0	S	S	S	S	S
10	V:V	9.32	0	0	S	S	S	S	S
11	I:V	9.15	0	0	S	S	S	S	S
12	I:L	9.06	0	0	S	S	S	S	S
13	D:H	8.82	0	0	S	S	S	S	S
14	N:S	8.73	0	0	S	S	S	S	S
15	S:S	8.22	0	0	S	S	S	S	S
16	K:Q	7.48	0	0	S	S	S	S	S
17	K:T	7.18	0	0	S	S	S	S	S
18	Q:T	7.10	0	0	S	S	S	S	S
19	P:W	6.98	0	0	S	S	S	S	S
20	H:H	6.42	0	0	S	S	S	S	S
21	D:K	6.38	0	0	S	S	S	S	S
22	F:L	6.35	0	0	S	S	S	S	S
23	D:T	5.47	0	0	S	S	S	S	S
24	F:F	5.46	0	0	S	S	S	S	S
25	D:Q	5.43	0	0	S	S	S	S	S
26	A:A	5.18	0	0	S	S	S	S	S
27	K:Y	5.08	0	0	S	S	S	S	S
28	K:S	5.03	0	0	S	S	S	S	S
29	R:T	4.95	0	0	S	S	S	S	S
30	N:N	4.87	0	0	S	S	S	S	S
31	D:S	4.81	0	0	S	S	S	S	S
32	L:M	4.73	0	0.00012	S	S	S	S	S
33	F:Y	4.55	0	0.00028	S	S	S	S	S
34	G:S	4.35	0.00002	0.00064	S	S	S	S	S
35	A:I	4.31	0.00002	0.00076	S	S	S	S	S
36	K:N	4.24	0.00002	0.00100	S	S	S	S	S
37	C:I	4.21	0.00002	0.00115	S	S	S	S	S
38	D:D	3.92	0.00008	0.00347	S	N	S	S	S
39	E:Q	3.90	0.00010	0.00367	S	N	S	S	S
40	E:T	3.51	0.00046	0.01453	S	N	S	N	S
41	F:G	3.46	0.00054	0.01716	S	N	S	N	S
42	A:L	3.39	0.00070	0.02129	S	N	S	N	S
43	A:Y	3.38	0.00072	0.02178	S	N	S	N	S
44	M:M	3.37	0.00076	0.02259	S	N	S	N	S
45	Q:S	3.34	0.00084	0.02503	S	N	S	N	S
46	L:V	3.31	0.00094	0.02739	S	N	S	N	S
47	M:V	3.29	0.00098	0.02858	S	N	S	N	S
48	P:Y	3.27	0.00106	0.03044	S	N	S	N	S
49	C:F	3.19	0.00142	0.03902	S	N	S	N	S
50	E:H	3.12	0.00178	0.04767	S	N	S	N	S
51	G:W	3.00	0.00268	0.06789	S	N	S	N	S
52	H:N	2.97	0.00302	0.07509	S	N	S	N	S
53	A:F	2.84	0.00444	0.09851	S	N	S	N	S
54	F:V	2.72	0.00650	0.14423	S	N	N	N	S
55	F:I	2.61	0.00908	0.19062	S	N	N	N	S
56	A:V	2.51	0.01214	0.24308	N	N	N	N	S
57	C:W	2.43	0.01494	0.28851	N	N	N	N	S
58	G:Y	2.40	0.01640	0.31148	N	N	N	N	S
59	F:W	2.32	0.00206	0.37510	N	N	N	N	S
60	G:G	2.26	0.02394	0.42423	N	N	N	N	S
61	L:W	2.23	0.02582	0.45096	N	N	N	N	S
62	Q:Q	2.20	0.02752	0.47496	N	N	N	N	S
63	G:V	2.15	0.03164	0.53167	N	N	N	N	S
					N	N	N	N	N
					:	:	:	:	:

Table 7: Unfavored amino acid pairs in antiparallel β -sheet bridges. See the caption of Table 4 for definitions of columns

Rank	Amino acid pair	STAR	p -value	Local FDR	IND	BF	B-H	EF	ST
210	L:T	-12.48	0	0	S	S	S	S	S
209	S:V	-8.57	0	0	S	S	S	S	S
208	I:N	-8.24	0	0	S	S	S	S	S
207	I:T	-8.19	0	0	S	S	S	S	S
206	L:S	-8.11	0	0	S	S	S	S	S
205	D:V	-8.06	0	0	S	S	S	S	S
204	G:K	-7.53	0	0	S	S	S	S	S
203	E:I	-7.46	0	0	S	S	S	S	S
202	I:S	-7.45	0	0	S	S	S	S	S
201	F:T	-7.40	0	0	S	S	S	S	S
200	T:Y	-7.00	0	0	S	S	S	S	S
199	K:V	-6.99	0	0	S	S	S	S	S
198	D:I	-6.88	0	0	S	S	S	S	S
197	T:V	-6.81	0	0	S	S	S	S	S
196	F:K	-6.73	0	0	S	S	S	S	S
195	A:K	-6.68	0	0	S	S	S	S	S
194	E:F	-6.51	0	0	S	S	S	S	S
193	I:K	-6.49	0	0	S	S	S	S	S
192	T:W	-6.39	0	0	S	S	S	S	S
191	F:N	-6.34	0	0	S	S	S	S	S
190	L:Q	-6.18	0	0	S	S	S	S	S
189	S:Y	-6.06	0	0	S	S	S	S	S
188	K:L	-6.03	0	0	S	S	S	S	S
187	D:L	-5.95	0	0	S	S	S	S	S
186	N:V	-5.75	0	0	S	S	S	S	S
185	L:N	-5.64	0	0	S	S	S	S	S
184	I:P	-5.62	0	0	S	S	S	S	S
183	E:L	-5.54	0	0	S	S	S	S	S
182	L:R	-5.44	0	0	S	S	S	S	S
181	D:F	-5.31	0	0	S	S	S	S	S
180	A:Q	-4.90	0	0	S	S	S	S	S
179	E:G	-4.83	0	0	S	S	S	S	S
178	I:Q	-4.68	0	0.00013	S	S	S	S	S
177	A:E	-4.67	0	0.00014	S	S	S	S	S
176	F:S	-4.33	0.00002	0.00062	S	S	S	S	S
175	D:W	-4.19	0.00002	0.00108	S	S	S	S	S
174	D:Y	-4.19	0.00002	0.00108	S	S	S	S	S
173	F:R	-4.14	0.00004	0.00137	S	S	S	S	S
172	R:V	-4.06	0.00004	0.00184	S	S	S	S	S
171	C:S	-4.02	0.00006	0.00212	S	N	S	S	S
170	C:T	-3.97	0.00008	0.00258	S	N	S	S	S
169	K:R	-3.95	0.00008	0.00288	S	N	S	S	S
168	C:Q	-3.93	0.00008	0.00308	S	N	S	S	S
167	H:L	-3.87	0.00010	0.00376	S	N	S	S	S
166	Q:V	-3.81	0.00014	0.00472	S	N	S	S	S
165	H:W	-3.70	0.00022	0.00716	S	N	S	S	S
164	G:R	-3.70	0.00022	0.00721	S	N	S	S	S
163	M:T	-3.68	0.00024	0.00760	S	N	S	S	S
162	C:E	-3.53	0.00040	0.01276	S	N	S	N	S
161	E:P	-3.45	0.00056	0.01710	S	N	S	N	S
160	G:T	-3.42	0.00062	0.01846	S	N	S	N	S
159	A:R	-3.33	0.00086	0.02493	S	N	S	N	S
158	I:R	-3.33	0.00088	0.02522	S	N	S	N	S
157	M:N	-3.32	0.00090	0.02611	S	N	S	N	S
156	H:V	-3.31	0.00092	0.02654	S	N	S	N	S
155	E:W	-3.24	0.00120	0.03318	S	N	S	N	S
154	E:V	-3.20	0.00138	0.03791	S	N	S	N	S
153	E:Y	-2.93	0.00334	0.08234	S	N	S	N	S
152	A:N	-2.93	0.00336	0.08274	S	N	S	N	S
151	N:Y	-2.86	0.00428	0.09234	S	N	S	N	S
150	A:T	-2.75	0.00588	0.13475	S	N	N	N	S
149	A:D	-2.63	0.00856	0.18603	S	N	N	N	S
148	F:Q	-2.61	0.00906	0.19529	S	N	N	N	S
147	M:Q	-2.45	0.01420	0.28603	N	N	N	N	S
146	A:C	-2.33	0.01954	0.37327	N	N	N	N	S
145	C:D	-2.24	0.02484 ²⁶	0.45520	N	N	N	N	S
144	G:H	-2.23	0.02574	0.46829	N	N	N	N	S
					N	N	N	N	N
					:	:	:	:	:
					:	:	:	:	:

Table 8: Summary of multiple testing results

Procedure	False positive rate	# of sig. favored pairs		# of sig. unfavored pairs	
		Parallel	Antiparallel	Parallel	Antiparallel
Individual test	Individual=0.01, FWER=0.88	39	55	38	63
Bonferroni	Individual=0.000048, FWER=0.01	23	37	25	39
Benjamin-Hochberg	FDR=0.01	31	53	35	60
Efron	Threshold of Local FDR=0.01	27	39	28	48
Storey	pFDR=0.01	42	63	40	67

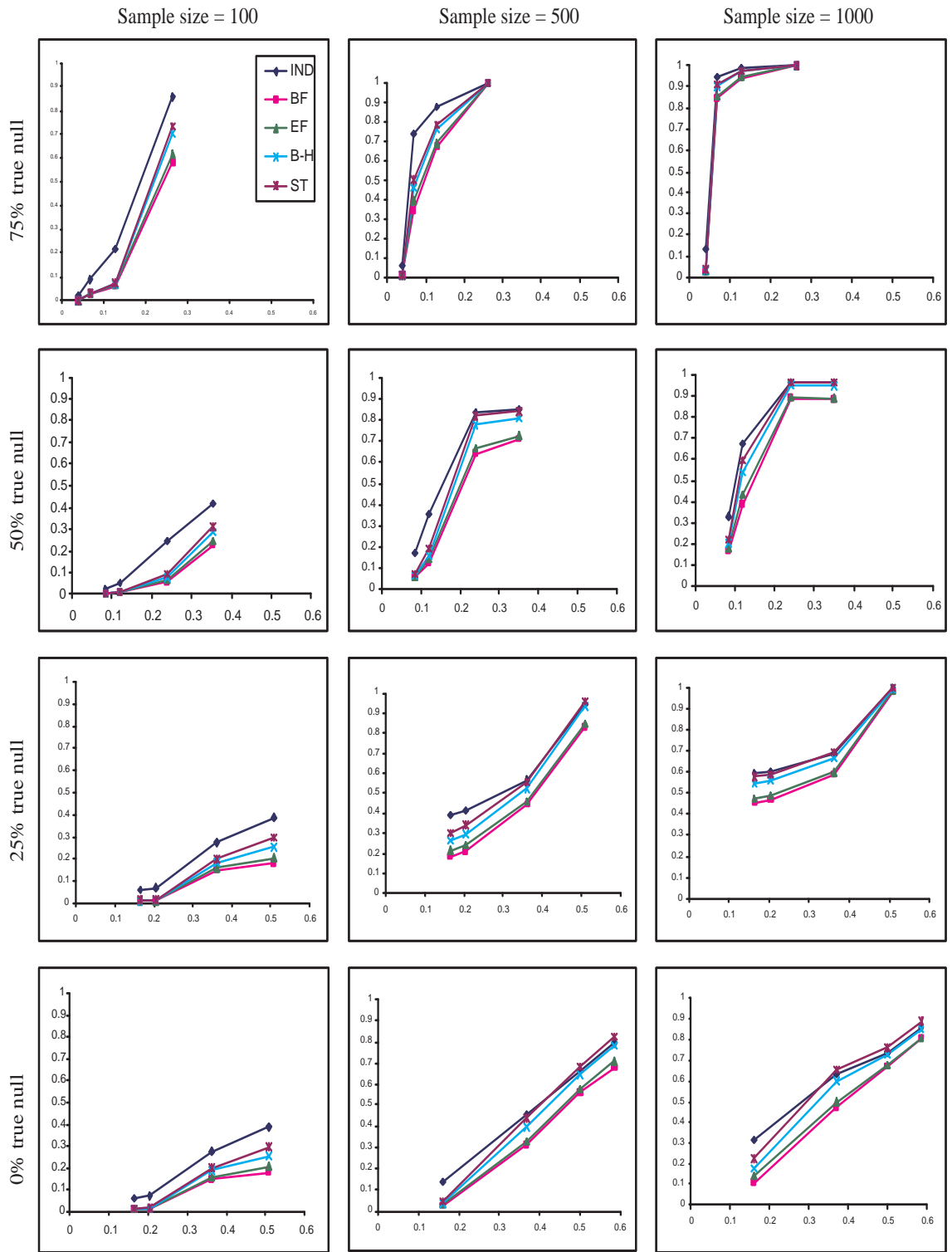


Figure 1: Simulation results of average empirical power.

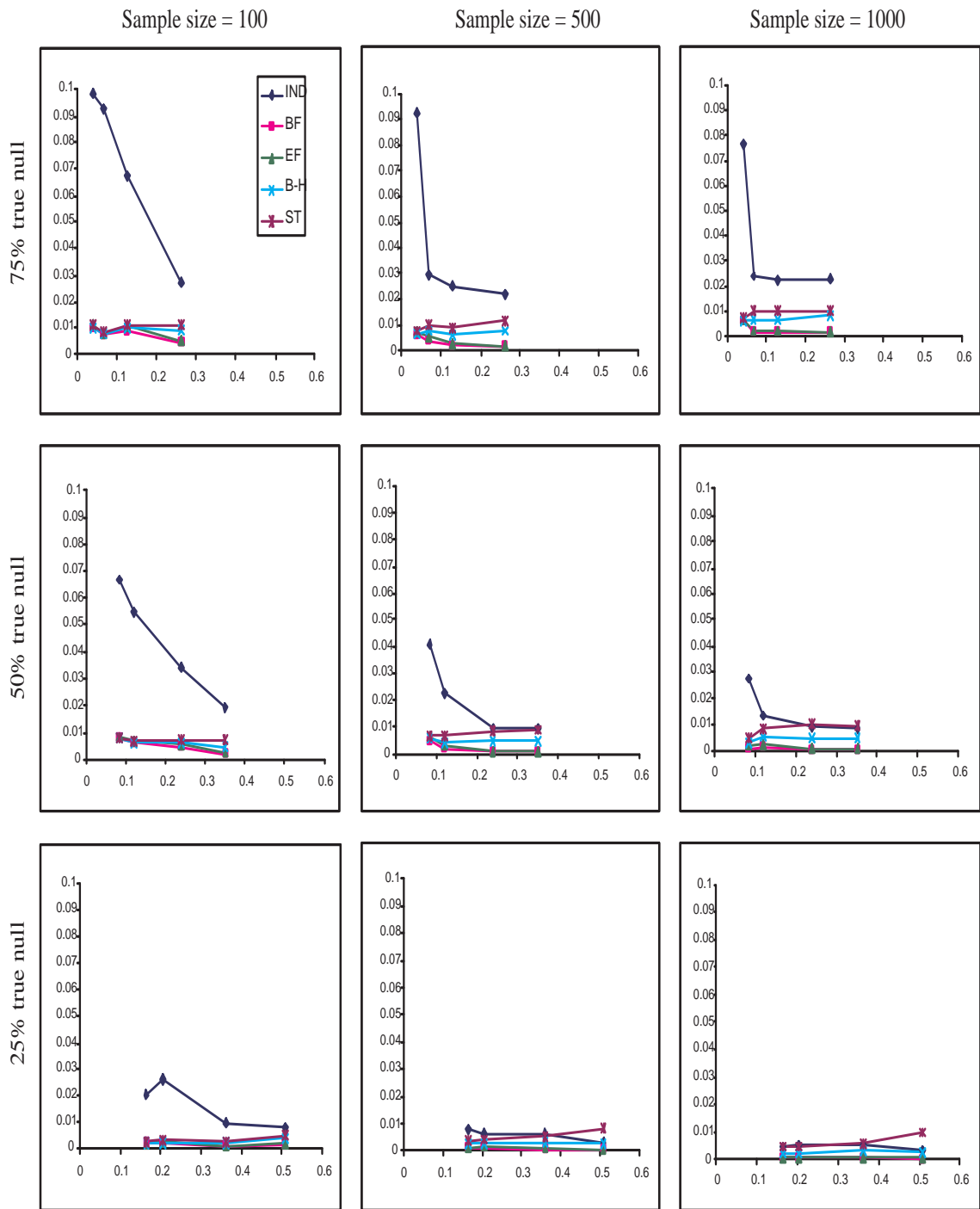


Figure 2: Simulation results of average empirical false discovery rate. Note that false discovery rate is 0 when the proportion of true null hypothesis is 0%.

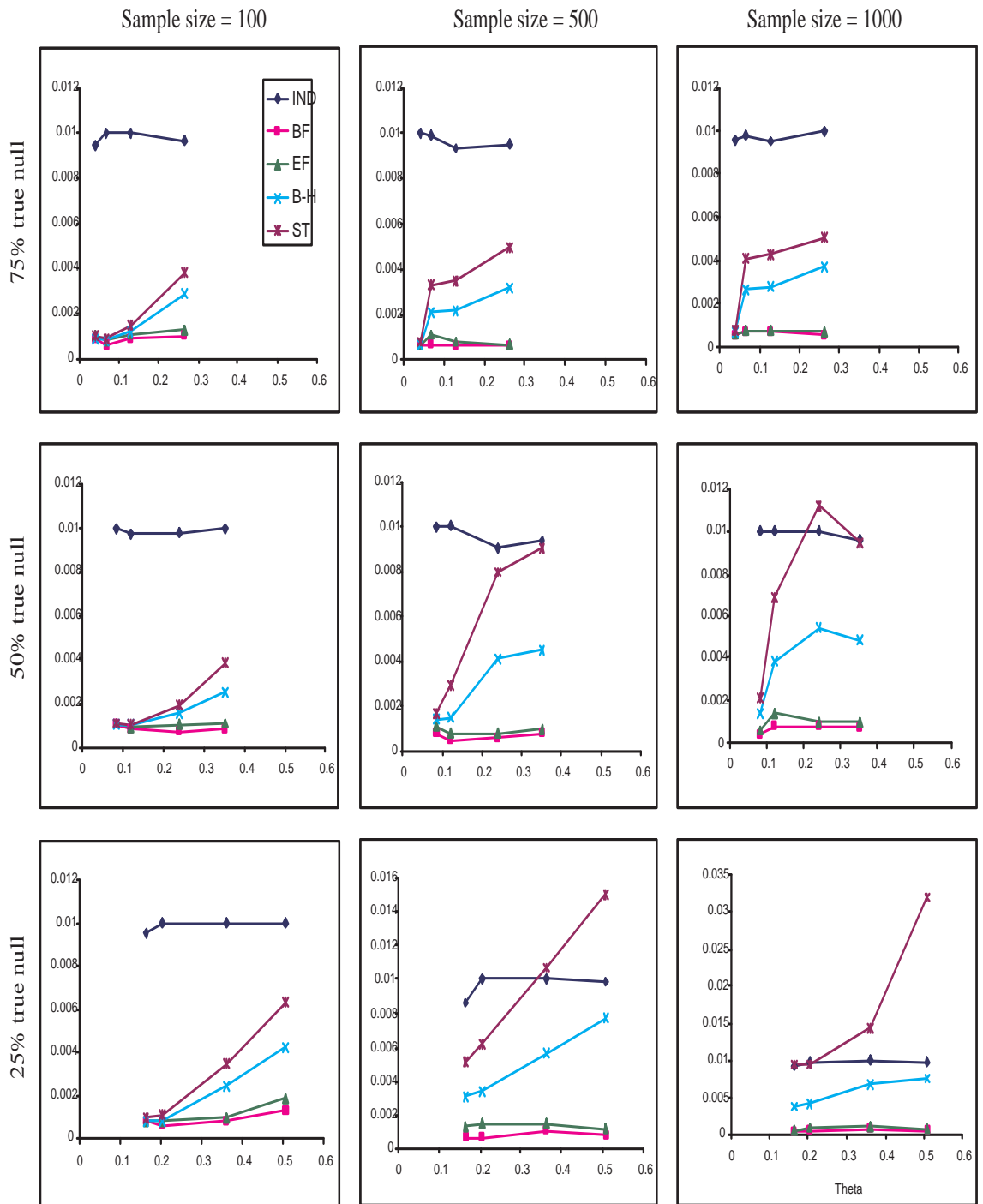


Figure 3: Simulation results of average empirical type I error: Note that type I error is not defined when the proportion of true null hypothesis is 0%.

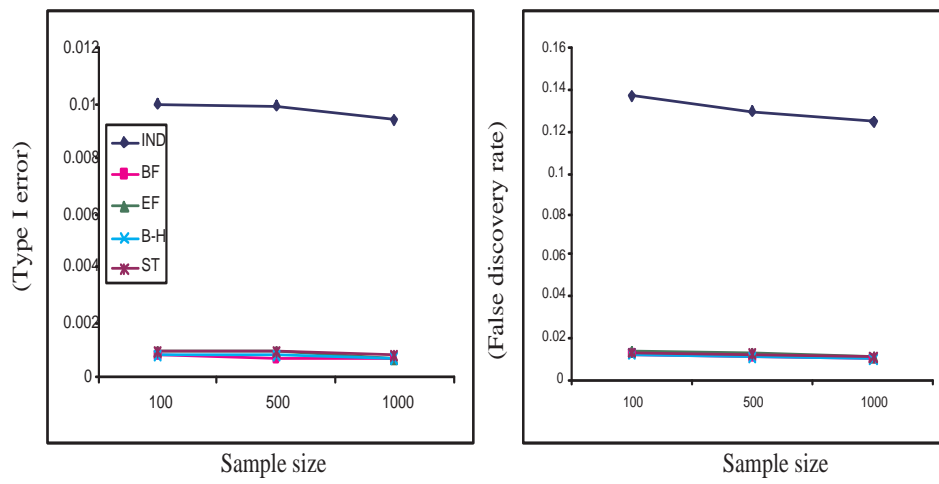


Figure 4: Average empirical type I error and false discovery rate when the proportion of true null hypothesis is 100%. Note that power is not defined in this case.