npg

Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix

J Li and L Ji

The Ministry of Education (MOE) Key Laboratory of Bioinformatics, Department of Automation, Tsinghua University, Beijing, People's Republic of China

Correlated multiple testing is widely performed in genetic research, particularly in multilocus analyses of complex diseases. Failure to control appropriately for the effect of multiple testing will either result in a flood of false-positive claims or in true hits being overlooked. Cheverud proposed the idea of adjusting correlated tests as if they were independent, according to an 'effective number' (M_{eff}) of independent tests. However, our experience has indicated that Cheverud's estimate of the M_{eff} is overly large and will lead to excessively conservative results. We propose a more accurate estimate of the M_{eff}, and design M_{eff}-based procedures to control the experiment-wise significant level and the false discovery rate. In an evaluation, based on both real and simulated data, the M_{eff}-based procedures were able to control the error rate accurately and consequently resulted in a power increase, especially in multilocus analyses. The results confirm that the M_{eff} is a useful concept in the error-rate control of correlated tests. With its efficiency and accuracy, the M_{eff} method provides an alternative to computationally intensive methods such as the permutation test.
Heredity (2005) **95**, 221–227. doi:10.1038/sj.hdy.6800717; published online 3 August 2005

Keywords: multilocus analysis; multiple testing; experiment-wise significant level; false discovery rate

Introduction

Complex diseases such as diabetes, schizophrenia and cancer lay a heavy burden on public health, with frequencies over 1%. As they are caused by interactions among genes, an increasing number of studies (Hoh *et al*, 2001; Nelson *et al*, 2001; Ritchie *et al*, 2001; Cordell and Clayton, 2002) have employed multilocus methods to detect the causative genes. Despite the abundance of such studies, only 16–30% of the reports of significant associations have been consistently replicated (Page *et al*, 2003). One reason for the low replication rate is the failure in controlling for the effect of multiple testing (Lander and Kruglyak, 1995).

Multilocus methods search loci in combinations, which inevitably tests a great number of hypotheses simultaneously: a k-locus analysis of n loci would have C_n^k locus-combinations to inspect. In addition to C_n^k being a large and cumbersome number, the tests are correlated with each other. For example, locus-combination AB is obviously correlated with combination BC for they share a mutual locus B. Even when loci are completely independent, their combinations are still correlated. These characteristics complicate the control of the error rate in multilocus analyses. Appropriate adjustment for multiple testing is very important to prevent a flood of false-positive claims and to prevent true associations being overlooked.

A traditional criterion for multiple testing is the experiment-wise significant level (EWSL, α_e), which is the probability that one or more hypotheses would be falsely rejected. To control the EWSL, the corresponding point-wise significant level (PWSL, α_p), which is the probability that an individual hypothesis would be falsely rejected, should be calculated appropriately. For independent tests, the Bonferroni (or Sidak) correction provides a simple and accurate control: $\alpha_p = 1-(1-\alpha_e)^{1/M} \approx \alpha_e/M$, where M is the number of tests. However, this adjustment is overly conservative for correlated tests. Lander and Botstein (1989) and Lander and Kruglyak (1995) offered analytical adjustments for correlated tests in single-locus analyses, but it is hard to apply them to multilocus analyses.

In practice, most researchers have therefore employed resampling-based methods to control the EWSL of correlated tests. Hoh *et al* (2001) and Ritchie *et al* (2001) used the permutation test; Nelson *et al* (2001) used the cross-validation test. Although resampling-based methods are accurate, they demand intensive computation, which constrains them to cases with a limited number of tests. Taking the permutation test as an example, Churchill and Doerge (1994) recommended at least 1000 shuffles to estimate a 0.05 EWSL and at least 10 000 shuffles to estimate a 0.01 EWSL.

An alternative criterion for multiple testing is the false discovery rate (FDR) (Benjamini and Hochberg, 1995), which is the expected proportion of falsely rejected hypotheses among all those rejected. When all hypotheses are true, the FDR is equal to the EWSL, but otherwise it is smaller. For independent tests, Benjamini and Hochberg (1995) and Benjamini and Liu (1999) proposed a step-up procedure and a step-down procedure to control the FDR. For correlated tests, Benjamini

Correspondence: *L Ji, The Ministry of Education (MOE) Key Laboratory of Bioinformatics, Department of Automation, Tsinghua University, Beijing 100084, People's Republic of China.*
E-mail: zc-sa@mail.tsinghua.edu.cn

and Yekutieli (2001) proposed a simple but highly conservative procedure. Storey (2002) and Storey and Tibshirani (2003) proposed a variation of the FDR: the positive false discovery rate (pFDR), which is the conditional FDR given that at least one hypothesis is rejected. These authors gave direct estimates of the pFDR for independent tests, and suggested resampling-based methods such as the permutation test for correlated tests. Correlation among tests is still a problem for FDR methods.

In 2001, Cheverud proposed a new idea to adjust correlated tests as though they were independent according to an effective number (M_{eff}) of independent tests. He also gave an estimation of the M_{eff} for single-locus analyses. In Nyholt (2004)'s evaluation, the M_{eff} method approximated the permutation test accurately, and was more than 100 times faster. It is highly desirable to apply this method to control the EWSL and FDR in multilocus analyses. However, our experience suggests that Cheverud's (2001) estimate of M_{eff} is overly large, especially when there are a substantial number of moderately correlated tests. This situation is, unfortunately, typical of multilocus analyses, leading to overly conservative results.

In this study, we propose a more accurate estimate of M_{eff}, and design M_{eff}-based procedures to control the EWSL and FDR. An assessment based on both real and simulated data shows that the procedures can control the error rate accurately in multilocus analyses.

Approach

The M_{eff} method

When loci are in linkage disequilibrium (LD), hypothesis tests in single-locus analyses are not independent. Cheverud (2001) proposed the method below to adjust the correlated tests:

Step 1: Calculate the correlation matrix for the loci.
Step 2: Estimate the effective number (M_{eff}) of independent tests from the eigenvalues of the correlation matrix with Equation (1), where M is the number of tests (equal to the number of loci in single-locus analyses) and λ_i ($i = 1, ..., M$) are the eigenvalues

$$\begin{cases} M_{eff} = 1 + (M-1)\left(1 - \dfrac{V_\lambda}{M}\right) \\ V_\lambda = \sum_{i=1}^{M} (\lambda_i - 1)^2 / (M-1) \end{cases} \quad (1)$$

Step 3: Adjust the test criteria as though there were M_{eff} independent tests with the Sidak (1967) correction below

$$\alpha_p = 1 - (1 - \alpha_e)^{1/M_{eff}} \quad (2)$$

Step 4: Test the association between genotypes and phenotypes locus by locus. If the p-value of any test is lower than α_p, the nonassociation hypothesis is rejected.

A new M_{eff} equation

The key step of the M_{eff} method is Step 2, estimating the M_{eff} from the eigenvalues. As Cheverud (2001) explained, Equation (1) considers two extreme situations. When the tests are completely independent, the eigen-

values are all equal to 1 and Equation (1) gives $M_{eff} = M$. When the tests are completely identical, the first eigenvalue is M, while all the others are zero, and Equation (1) gives $M_{eff} = 1$. Although Equation (1) gives correct estimates at the two extremes, it overestimates the M_{eff} in other situations. If the M tests are composed of c ($1 \leqslant c \leqslant M$) copies of M/c independent tests, the eigenvalues are Sequence (3)

$$\overbrace{c, c, \ldots, c}^{M/c}, \overbrace{0, 0, \ldots, 0}^{(c-1)M/c} \quad (3)$$

According to Equation (1), $M_{eff} = M + 1 - c$, but actually there are only M/c independent tests. The ratio r is

$$r = \frac{M+1-c}{M/c} = \frac{c(M+1-c)}{M} \geqslant 1, 1 \leqslant c \leqslant M \quad (4)$$

In multilocus analyses, M is large and c is moderate, so r tends to be considerably larger than 1.

Intuitively, c identical tests will result in an eigenvalue $\lambda_i = c$, while a test partially correlated with others will result in an eigenvalue $0 < \lambda_i < 1$. Thus, an eigenvalue can be decomposed into two parts: the integral part and the nonintegral part. The integral part represents identical tests that should be counted as one in the M_{eff}. The nonintegral part represents a partially correlated test that should be counted as a fractional number between 0 and 1. From this intuitive consideration, we propose Equation (5) to estimate the M_{eff}:

$$\begin{cases} M_{eff} = \sum_{i=1}^{M} f(|\lambda_i|) \\ f(x) = I(x \geqslant 1) + (x - \lfloor x \rfloor), x \geqslant 0 \end{cases} \quad (5)$$

where $I(x \geqslant 1)$ is the indicator function, which gives 1 when $x \geqslant 1$ and gives 0 otherwise, and $\lfloor x \rfloor$ is the floor function, which gives the largest integer less than or equal to x. Thus, completely correlated tests will be counted as $I(x \geqslant 1)$ and partially correlated tests will be counted as $x - \lfloor x \rfloor$. Equation (5) gives the right M_{eff} for integral eigenvalue Sequence (6), where t is the true number of independent tests and $c_1, c_2, ..., c_t$ are integers

$$\overbrace{c_1, c_2, \ldots, c_t}^{t}, \overbrace{0, 0, \ldots, 0}^{M-t} \quad (6)$$

Multilocus analyses

By regarding locus-combinations as 'generalized loci' (GLs), we can apply the M_{eff} method to multilocus analyses. For example, if two loci have g_1 and g_2 genotypes, respectively, their combination can be regarded as a new 'locus' with $g_1 g_2$ genotypes. Consequently, the calculation of the correlation coefficients in Step 1 needs adaptation. As GLs are multinomial variables, neither the Pearson correlation (used by Cheverud, 2001) nor the LD measure (used by Nyholt, 2004) is suitable.

Similar to the Pearson correlation for numerical variables, the correlation for multinomial variables can be defined as in Equation (7a), but the variance (Var) and the covariance (Cov) should be measured differently. Let X_1 and X_2 denote two GLs, that is, two multinomial variables. Their variance and covariance can be measured with the entropy and mutual information (Cover

and Thomas, 1991), respectively, as in Equation (7b) and (7c).

$$\text{Cor}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} \tag{7a}$$

$$\text{Var}(X_i) = E[-\log P(X_i)], \ i = 1, 2 \tag{7b}$$

$$\text{Cov}(X_1, X_2) = E\left[\log \frac{P(X_1 X_2)}{P(X_1)P(X_2)}\right] \tag{7c}$$

where $P(X_i)$ is the probability function of X_i and $E[x]$ is the expectation of a random variable x.

Besides its compatibility with both single- and multilocus analysis, this correlation coefficient has a property:

$$\begin{aligned} \text{Cor}(AB, BC) &= \frac{\text{Var}(B)}{\sqrt{\text{Var}(AB)\text{Var}(BC)}} \\ \text{Var}(AB) &= \text{Var}(A) + \text{Var}(B) \\ \text{Var}(BC) &= \text{Var}(B) + \text{Var}(C) \end{aligned} \tag{8}$$

where A, B and C are independent loci or GLs.

An M_{eff}-based FDR procedure

Let R denote the number of rejected hypotheses and V the number of those falsely rejected. Benjamini and Hochberg (1995) defined the FDR as

$$\text{FDR} = E\left[\frac{V}{R}I(R > 0)\right]$$

where $I(R > 0)$ is the indicator function, which gives 1 when $R > 0$ and gives 0 otherwise. The FDR is equal to the EWSL when all null hypotheses are true and is smaller otherwise. They also proposed the following step-up procedure to control the FDR for independent tests. Let $p_{(1)} \leqslant \cdots \leqslant p_{(M)}$ be the ordered p-values of M independent hypotheses $H_{(1)}, \ldots, H_{(M)}$. To control the FDR at level q (usually 5%), hypotheses $H_{(1)}, \ldots, H_{(k)}$ should be rejected, where

$$k = \max_{1 \leqslant i \leqslant M}\left\{i : p_{(i)} \leqslant \frac{i}{M}q\right\}$$

The ordered p-values are compared with an arithmetic sequence from q/M to q. The start of the sequence, q/M, ensures that the FDR is equal to the EWSL when all the

hypotheses are true. The end of the sequence, q, ensures that the FDR is controlled at q when all the hypotheses are rejected. To involve the effect of correlation, we can replace the start q/M with q/M_{eff}, keep the end as q, and consequently replace the step q/M with $(q-q/M_{eff})/(M-1)$. The M_{eff}-based step-up procedure is to reject $H_{(1)}, \ldots, H_{(k)}$, where

$$k = \max_{1 \leqslant i \leqslant M}\left\{i : p_{(i)} \leqslant \frac{q}{M_{eff}} + \frac{i-1}{M-1}\left(q - \frac{q}{M_{eff}}\right)\right\}$$

Performance

Single-locus analyses

The M_{eff} method has previously been applied to two single-locus analyses. Cheverud (2001) applied it to the *LG/J* and *SM/J* mouse data (Cheverud *et al*, 2001), and Nyholt (2004) applied it to the *angiotensin-I-converting enzyme* (*ACE*) gene data (Keavney *et al*, 1998). In these analyses, Cheverud (2001) and Nyholt (2004) derived two eigenvalue sequences from the two data sets, respectively. In this study, we calculated different M_{eff}'s with both Equations (1) and (5) from the two eigenvalue sequences, and compared the results. As shown in Table 1, the two equations gave similar M_{eff}'s, so they should have similar performance on the two data sets. Setting $\alpha_e = 5\%$ for the Keavney data set, Nyholt (2004) obtained $\alpha_p = 0.015$ from a permutation test with 50 000 shuffles, and obtained $\alpha_p = 0.011$ with the $M_{eff} = 4.59$ from Equation (1). Using $M_{eff} = 4.01$ from Equation (5), we obtained $\alpha_p = 0.013$, a value closer to the permutation test's 0.015 than the value of 0.011 obtained from Equation (1).

Multilocus analyses

To compare the performance of Equations (1) and (5) in multilocus analyses, we used four sub-data sets extracted from a schizophrenia data set (Xu *et al*, 2004). The schizophrenia data set concerns 12 genes of the dopamine pathway, *COMT* (7), *ALDH*3 (11), *ADH* (14), *TYR* (1), *TPO* (2), *MAO* (3), *DAO* (2), *AOX*1 (2), *DDC* (4), *HDG* (1), *DBH* (2) and *GSTZ*1 (1), with 83 cases and 108 controls. (The integers in the parentheses are the numbers of SNPs genotyped for each gene.) To investigate the performance on correlated loci, we extracted the SNPs of gene *COMT*, *ALDH*3 and *ADH* as three sub-data sets COMT, ALDH3 and ADH, respectively. The three genes were reported as potential effective SNPs by Xu

Table 1 Comparison between Equations (1) and (5) on two eigenvalue sequences

Eigenvalue sequence	Permutation		Equation (1)		Equation (5)	
	M_{eff}	α_p	M_{eff}	α_p	M_{eff}	α_p
Cheverud 3.63, 1.79, 1.13, 0.47, 0.37, 0.25, 0.18, 0.17	Not given by Cheverud		6.73	0.76%	5.99	0.85%
Nyholt 7.84, 1.60, 0.24, 0.21, 0.08, 0.03, 0.01, 0, 0, 0	3.39 (50 000 shuffles)	1.5%	4.59	1.1%	4.01	1.3%

Cheverud (2001)'s eigenvalue sequence was derived from the *LG/J* and *SM/J* mouse data (Cheverud *et al*, 2001) with the Pearson correlation. Nyholt (2004)'s sequence was derived from the *ACE* gene data (Keavney *et al*, 1998) with the LD measure. The EWSL was set at $\alpha_e = 5\%$.

et al (2004), and they have enough SNPs (more than five) to ensure the multiplicity of tests in the analyses. To investigate the performance on independent loci, we selected one SNP per gene as the fourth sub-data set and called it DOP (short for dopamine). For each data set, we performed single-, two- and three-locus analyses.

We used the permutation test as the golden standard to evaluate the performance. In each permutation shuffle, we applied the M_{eff} method to control the EWSL (α_e). As a result, we obtained a decision of whether or not to reject the nonassociation hypothesis. After repeating this procedure 3000 times, we counted the proportion of rejection (P_r). Ideally, P_r should be equal to the preset α_e. The association between a GL and the phenotype was tested with a combination of the χ^2 test and the Fisher's exact test (When Cochran (1954)'s condition was satisfied, the association was tested with the χ^2 test. When the condition was not satisfied, it was tested with the Fisher's exact test. Cochran's condition: in a contingency table, every cell has an expected value no less than 1, and 80% of the cells have expected values of no less than 5. This algorithm was implemented in ACM algorithm 643 (Clarkson *et al*, 1993).)

Figure 1 shows the comparison between P_r and α_e. In single-locus analyses, both Equations (1) and (5) controlled P_r at α_e. In two- and three-locus analyses, the M_{eff}'s obtained from Equation (1) were much greater

than those obtained from the permutation test (see Table 2), and consequently the P_r of Equation (1) was much lower than α_e. In contrast, the P_r of Equation (5) approximated α_e fairly well in most cases.

Effects on FDR

We performed two-locus analyses on simulated case–control data to evaluate the effects of the M_{eff} on the FDR. The simulated data contains 10 genomic regions with five observed biallelic SNPs in each region. All SNPs are in the Hardy–Weinberg equilibrium and the minor alleles have the same frequency. SNPs in different regions are in linkage equilibrium, and SNPs in the same region are in LD with $r^2 = 0.8$ (Hill and Robertson, 1968) between two successive SNPs. The disease is caused by the epistatic interaction of two unobserved loci, which hide in the first region and the second region, respectively. The LD (r^2) between a causative loci and its nearest marker SNP is 0.8. As neither of the causative loci has a marginal effect, a two-locus analysis is necessary to detect the epistatic interaction. Consequently, there are $C_{50}^2 = 1225$ SNP pairs to inspect, and only the 25 pairs composed of the loci from the first and second regions are associated with the disease. The data were simulated from the six epistasis models in Table 3 (Ritchie *et al*,
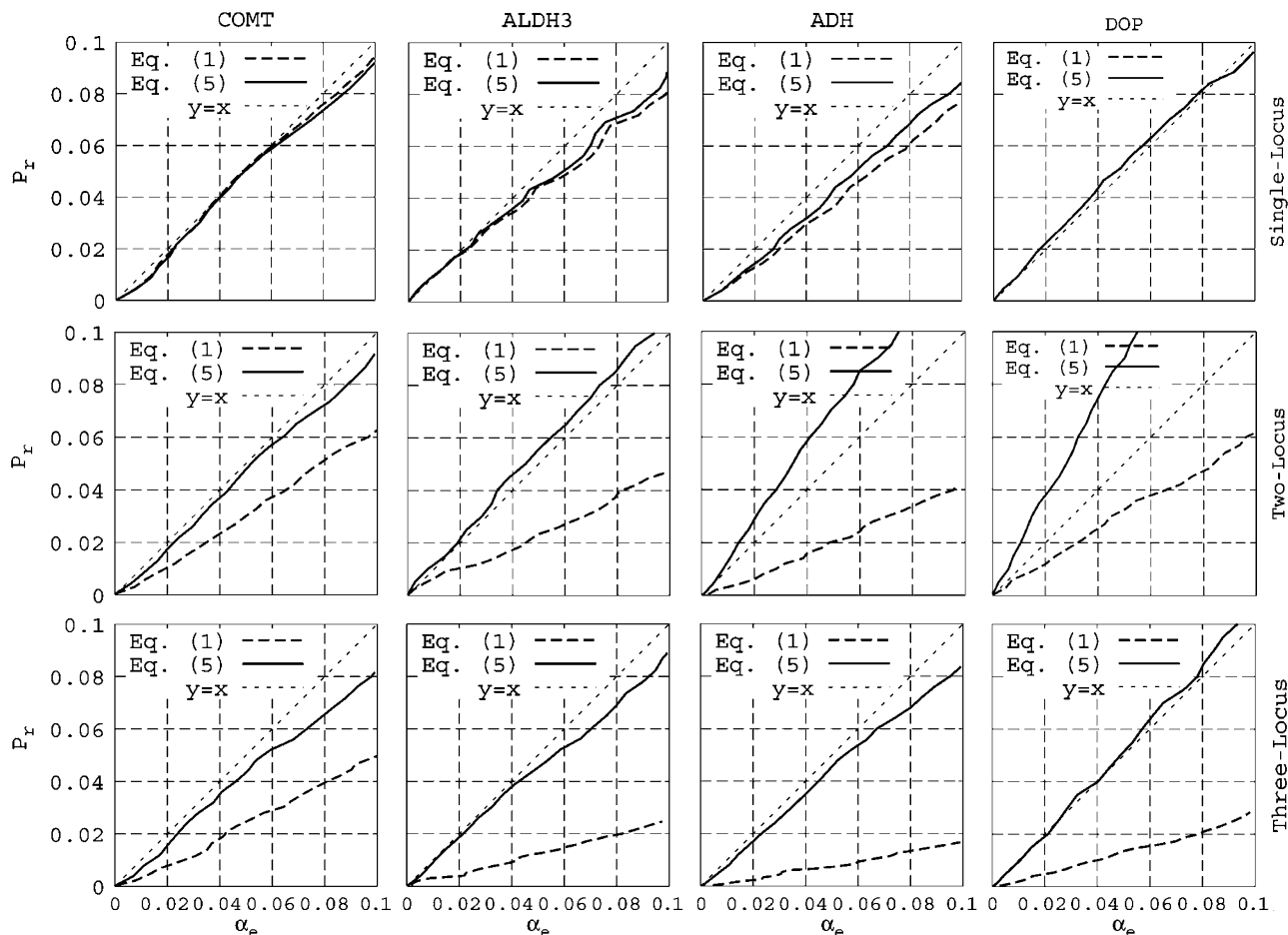

Figure 1 Proportion of rejection *vs* EWSL. The EWSL (α_e) was set at different levels, and then the proportion of rejection (P_r) was estimated with a permutation test with 3000 shuffles. Ideally, P_r should be equal to α_e. The dashed lines are the result of the M_{eff} from Equation (1), and the solid lines are the result of the M_{eff} from Equation (5). All correlation matrices were calculated with Equation (7).

Table 2 Comparison between Equations (1) and (5) on the schizophrenia data

Analyses	Permutation		Equation (1)		Equation (5)	
	M_{eff}	α_p (%)	M_{eff}	α_p (%)	M_{eff}	α_p (%)
COMT						
Single locus	7.01	0.729	6.84	0.747	7.00	0.730
Single, Pearson*	7.01	0.729	6.17	0.828	6.00	0.851
Two locus	10.71	0.478	16.86	0.304	11.09	0.461
Three locus	10.77	0.475	22.17	0.231	12.27	0.417
ALDH3						
Single locus	8.47	0.603	10.53	0.486	10.01	0.511
Single, Pearson*	8.47	0.603	9.95	0.514	9.00	0.568
Two locus	20.28	0.253	44.28	0.116	18.19	0.282
Three locus	25.48	0.201	111.80	0.0459	27.87	0.184
ADH						
Single locus	10.45	0.490	13.43	0.381	12.05	0.425
Single, Pearson*	10.45	0.490	12.36	0.414	10.00	0.512
Two locus	31.77	0.162	76.42	0.0671	21.78	0.235
Three locus	45.9	0.114	264.97	0.0195	49.77	0.103
DOP						
Single locus	13.01	0.394	11.99	0.427	12.02	0.426
Single, Pearson*	13.01	0.394	11.87	0.431	12.00	0.427
Two locus	35.01	0.146	59.63	0.0856	19.42	0.264
Three locus	48.32	0.106	179.40	0.0286	47.12	0.109

Theoretically, the M_{eff} of Equation (5) should be an integer. However, we obtained real numbers because of the computational errors in the calculation of eigenvalues. In the four analyses marked by asterisks (*), the correlation matrices were calculated with the Pearson correlation. In all the other analyses, the correlation matrices were calculated with Equation (7). The EWSL was set at $\alpha_e = 5\%$.

Table 3 Epistasis models of disease

	Model 1			Model 2			Model 3		
	BB	Bb	Bb	BB	Bb	bb	BB	Bb	bb
AA	0	1	0	0	0	2	8	7	5
Aa	1	0	1	0	1	0	10	0	10
aa	0	1	0	2	0	0	3	10	4
		$P(A) = P(B) = 0.5$			$P(A) = P(B) = 0.5$			$P(A) = P(B) = 0.25$	
	Model 4			Model 5			Model 6		
	BB	Bb	Bb	BB	Bb	bb	BB	Bb	bb
AA	0	1	9	7	5	2	9	0.1	2
Aa	4	1	8	5	9	1	8	7	0.5
aa	7	9	3	2	1	3	0.3	0.7	2
		$P(A) = P(B) = 0.25$			$P(A) = P(B) = 0.1$			$P(A) = P(B) = 0.1$	

The table describes the epistasis models used by Ritchie *et al* (2003), where $P(A)$ and $P(B)$ are the minor allele frequencies, and the 3×3 tables are the relative penetrances. These models have little or no marginal effects at each locus. The real penetrances can be calculated by using the population penetrance. The population penetrance was set at 1% in the simulation study because this is the penetrance of complex diseases (Hoh and Ott, 2003).

2003). For each model, 10 000 data sets were generated with 200 cases and 200 controls in each of them.

Both Benjamini and Hochberg (1995)'s direct step-up procedure and the new M_{eff}-based step-up procedure were applied to control the FDR. Although Benjamini and Yekutieli (2001) designed an FDR procedure for correlated multiple testing, we did not employ it because it is much more conservative than the direct step-up procedure. The M_{eff} was estimated with Equation (5). As shown in Figure 2, the M_{eff}-based step-up procedure controlled the FDR more accurately than the direct method, and was considerably more powerful.

Discussion and conclusion

In 2001, Cheverud proposed the idea of adjusting correlated tests as if they were independent according to the effective number (M_{eff}) of independent tests. Piepho (2001) has also proposed a quick method for computing approximate threshold levels that control the genome-wise type I error rate of tests for quantitative trait locus detection in interval mapping and composite interval mapping. In the multilocus analyses presented in this paper, our extended concept of M_{eff} succeeded in controlling both the EWSL and the FDR. This result
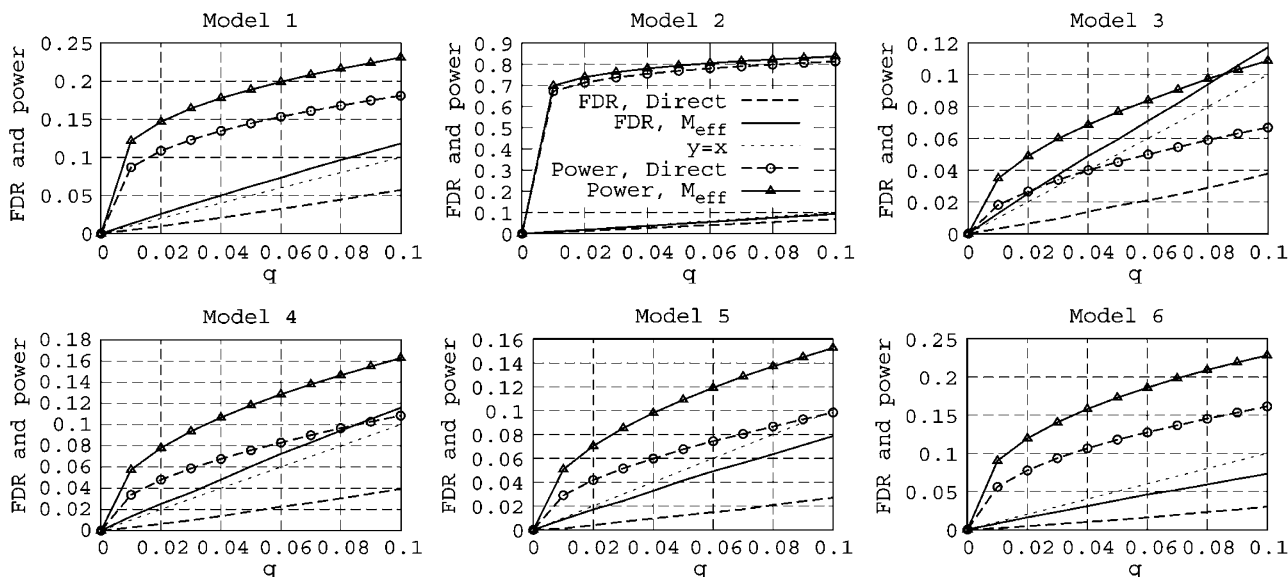
Figure 2 Comparison between the direct step-up procedure and the M_{eff}-based step-up procedure. The FDR and power were estimated from 10 000 data sets, each with 200 cases and 200 controls. Ideally, the empirical FDR should be equal to q, which is the FDR level expected to be controlled at.

supports the use of M_{eff} in error-rate control of correlated multiple testing. As a data-driven algorithm based on a correlation matrix, the M_{eff} method can be applied to a wide range of problems. Nowadays, dense and linked markers are widely used in genetic study, and it is hard to model their genetic behavior precisely, especially in multilocus analyses. The diversity of populations also prohibits models obtained from one population from being applied to another. Avoiding the error of an imprecise model, the data-driven M_{eff} method is suitable for different situations. In addition to the data-driven property, the M_{eff} can be calculated for various types of data by generalizing the definition of correlation coefficient. With Equation (7), it can be applied to categorical data. With the Pearson correlation, it can be applied to numerical data such as microarray data. With King and Chinchilli (2001)'s generalized correlation coefficient, it can be applied to a mixture of numerical and categorical data.

Cheverud (2001) proposed Equation (1) to estimate the M_{eff}. As he explained, Equation (1) considers two extreme situations: hypothesis tests are completely independent or they are completely identical. However, Equation (4) shows that Equation (1) tends to overestimate the M_{eff} when there are a large number of moderately correlated tests. Unfortunately, in multilocus analyses, this situation is usually encountered, and Equation (1) will lead to conservative reports. In this paper, we propose Equation (5) to estimate the M_{eff} more accurately. It considers a more general situation than the two extremes Equation (1) concerns, where tests are composed of copies of independent tests. Therefore, it outperforms Equation (1) in multilocus analyses. Figure 1 clearly shows that both Equations (1) and (5) perform excellently in single-locus analyses, but only the latter is effective in multilocus analyses. The higher performance of Equation (5) over Equation (1) is not because it better fits the correlation coefficient defined by Equation (7). As

Equation (4) has shown, no matter what correlation coefficient is employed to calculate the correlation matrix, Equation (1) will overestimate the M_{eff} for eigenvalue Sequence (3), while Equation (5) will give an appropriate estimate. We also used the Pearson correlation in single-locus analyses, but as shown in Table 2, its performance was similar to that of Equation (7).

If n tests need adjustment, the computation complexity of a permutation test is $O(stn)$, where s is the number of shuffles and t is the time needed to test a hypothesis. The computation cost for the M_{eff} method concentrates on eigenvalues' calculation whose complexity is $O(mn^3)$, where m is the time needed to multiply two real numbers. Theoretically, when n is large, the permutation test is more efficient than the M_{eff} method. However, because m is tiny in comparison with st, the M_{eff} method is much faster than the permutation test until n is very large. The largest n we encountered in this study was 1225 in Section Effects on FDR. It took about 10 000 s to repeat the calculation of the 1225 p-values 10 000 times, and took about 3 s to solve the eigenvalues with the numeric package *Lapack*. (The computation time was measured on a 2.8 GHz Pentium with Linux 2.6.10.) When there are about 1000 tests, the M_{eff} method is at least 1000 times faster than the permutation test. This would meet the need of a genome scan with about 1000 SNPs per chromosome. As SNPs in different chromosomes are in linkage equilibrium, the correlation matrix of a single-locus analysis is diagonally blocked. We can calculate the M_{eff} for each chromosome individually and then sum them up as the M_{eff} for the scan.

Freed from the heavy computation burden, statisticians can realize many sophisticated multilocus methods. For example, we can split a classification tree's nodes according to the EWSL or the FDR rather than the PWSL. In a tree's growth procedure, a node splits when the p-value of the optimal variable exceeds a threshold. The more the variables are, the easier it is for the optimal

variable to exceed the threshold, and the more easily the tree overfits. Zhang and Bonney (2000) considered this problem in splitting, but they arbitrarily set the PWSL < 0.001 as the threshold. With the M_{eff} method, the threshold can be estimated efficiently. If the threshold is calculated with the permutation test, the computation will be very demanding because every node needs thousands of shuffles.

Multiple testing is not only important in data analysis but also in experiment design. Neglecting the issue would lead to an underpowered design, while over-emphasizing it would waste resources. Owing to the expense of genetic experiments, it is highly desirable to apply the M_{eff} method to experiment designs. By calculating the M_{eff} from a pilot study, we can estimate the appropriate PWSL and the sample size for an experiment involving multiple testing.

Acknowledgements

References

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B (Methodological)* **57**: 289–300.

Benjamini Y, Liu W (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *J Stat Plan Infer* **82**: 163.

Benjamini Y, Yekutieli D (2001). The control of the false discovery rate in multiple testing under dependency. *Ann Stat* **29**: 1165–1188.

Cheverud JM (2001). A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* **87**: 52–58.

Cheverud JM, Vaughn TT, Pletscher LS, Peripato AC, Adams ES, Erikson CF et al (2001). Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. *Mamm Genome* **12**: 3–12.

Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.

Clarkson DB, Fan Y-a, Joe H (1993). A remark on algorithm 643: FEXACT: an algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *ACM Trans Math Software (TOMS)* **19**: 484–488.

Cochran W (1954). Some methods of strengthening the common χ^2 tests. *Biometrics* **10**: 417–451.

Cordell HJ, Clayton DG (2002). A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* **70**: 124–141.

Cover TM, Thomas JA (1991). *Elements of Information Theory.* Wiley-Interscience: New York.

Hill WG, Robertson A (1968). Linkage disequilibrium in finite populations. *Theoret Appl Genet* **38**: 226–231.

Hoh J, Ott J (2003). Mathematical multi-locus approaches to localizing complex human trait genes. *Nat Rev Genet* **4**: 701–709.

Hoh J, Wille A, Ott J (2001). Trimming, weighting, and grouping SNPs in human case–control association studies. *Genome Res* **11**: 2115–2119.

Keavney B, McKenzie CA, Connell JM, Julier C, Ratcliffe PJ, Sobel E et al (1998). Measured haplotype analysis of the angiotensin-I converting enzyme gene. *Hum Mol Genet* **7**: 1745–1751.

King TS, Chinchilli VM (2001). A generalized concordance correlation coefficient for continuous and categorical data. *Stat Med* **20**: 2131–2147.

Lander E, Kruglyak L (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**: 241–247.

Lander ES, Botstein D (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.

Nelson MR, Kardia SL, Ferrell RE, Sing CF (2001). A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. *Genome Res* **11**: 458–470.

Nyholt DR (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* **74**: 765–769.

Page GP, George V, Go RC, Page PZ, Allison DB (2003). 'Are we there yet'? Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *Am J Hum Genet* **73**: 711–719.

Piepho H-P (2001). A quick method for computing approximate thresholds for quantitative trait loci detection. *Genetics* **157**: 425–432.

Ritchie MD, Hahn LW, Moore JH (2003). Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genet Epidemiol* **24**: 150–157.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF et al (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* **69**: 138–147.

Sidak Z (1967). Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc* **62**: 626–633.

Storey JD (2002). A direct approach to false discovery rates. *J Roy Stat Soc Ser B* **64**: 479–498.

Storey JD, Tibshirani R (2003). Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100**: 9440–9445.

Xu Q, Jia YB, Zhang BY, Zou K, Tao YB, Wang YP et al (2004). Association study of an SNP combination pattern in the dopaminergic pathway in paranoid schizophrenia: a novel strategy for complex disorders. *Mol Psychiatry* **9**: 510–521.

Zhang H, Bonney G (2000). Use of classification trees for association studies. *Genet Epidemiol* **19**: 323–332.