

RESEARCH ARTICLE

Open Access

Generalized shrinkage F -like statistics for testing an interaction term in gene expression analysis in the presence of heteroscedasticity

Jie Yang^{1*}, George Casella^{2,4} and Lauren M McIntyre^{2,3,4}

Abstract

Background: Many analyses of gene expression data involve hypothesis tests of an interaction term between two fixed effects, typically tested using a residual variance. In expression studies, the issue of variance heteroscedasticity has received much attention, and previous work has focused on either between-gene or within-gene heteroscedasticity. However, in a single experiment, heteroscedasticity may exist both within and between genes. Here we develop flexible shrinkage error estimators considering both between-gene and within-gene heteroscedasticity and use them to construct F -like test statistics for testing interactions, with cutoff values obtained by permutation. These permutation tests are complicated, and several permutation tests are investigated here.

Results: Our proposed test statistics are compared with other existing shrinkage-type test statistics through extensive simulation studies and a real data example. The results show that the choice of permutation procedures has dramatically more influence on detection power than the choice of F or F -like test statistics. When both types of gene heteroscedasticity exist, our proposed test statistics can control preselected type-I errors and are more powerful. Raw data permutation is not valid in this setting. Whether unrestricted or restricted residual permutation should be used depends on the specific type of test statistic.

Conclusions: The F -like test statistic that uses the proposed flexible shrinkage error estimator considering both types of gene heteroscedasticity and unrestricted residual permutation can provide a statistically valid and powerful test. Therefore, we recommended that it should always applied in the analysis of real gene expression data analysis to test an interaction term.

Background

The regulation of gene expression starts when a cell's DNA is transcribed into mRNA. The simultaneous expression profiles of many genes under different circumstances can provide insight into physiological processes. Using modern technologies in gene expression experiments such as oligonucleotide arrays [1], and cDNA spotted arrays [2], many scientists have made novel discoveries about complex biological processes of yeast [3,4], drosophila [5], mice [6], humans [7], and other species. Recently one such study also included RNA-seq [8]. Statistical methodologies and issues

involved in microarray data analysis have been widely reviewed [9-12], and it is expected that many of the same issues will need to be addressed with RNA-seq.

The analysis of variance (ANOVA) model is a popular statistical modeling method for the analysis of microarrays. Since its introduction by Kerr *et al.* [13], it has been extensively examined for use in this setting [14-21]. Kerr *et al.* constructed an ANOVA model that included the gene effect as a fixed effect. This model assumes identically and independently distributed residual errors across genes. The advantage of this model is that the large number of genes involved in a microarray experiment results in huge degrees of freedom for the error estimate, which can lead to a very powerful test. However, the common assumption of homoscedasticity may not hold true in this setting [22]. One alternative is

* Correspondence: jie.yang@stonybrook.edu

¹Department of Preventive Medicine, Stony Brook University, Stony Brook, NY 11794, USA

Full list of author information is available at the end of the article

to use an ANOVA model for each gene, but the resulting test statistics from gene-specific models may have limited power because the biological sample size for each gene in a microarray experiment is usually small.

To address this problem of limited power, researchers have proposed other methods for obtaining more information across genes, ranging from a simple equal-weighted average of a gene-specific error estimate and the global average of all gene-specific error estimates (F_2 statistic proposed by Wu *et al.* [19] to empirical Bayesian modeling of all gene-specific errors [23-26]. Other variations [27-29] used different variance modeling strategies to address the heteroscedasticity problem, but no clear winner has emerged [30]. Huang and Liu [31] extended the test statistics proposed by Cui *et al.* [28] by assuming a normal distribution on the mean and then deriving an empirical Bayes likelihood ratio test. The resulting test statistic shrinks both the mean and variances.

In addition to the problem of *between-gene* heteroscedasticity, we must also be concerned with *within-gene* heteroscedasticity. For example, in the study of simple differential gene expression between a treatment group and a control group, the variance in the treatment arm may differ from that in the control arm. Some approaches to this problem include a general Bayesian framework to model heteroscedastic error in a single generalized linear mixed model setting [32] and a structural model placed on the error variances specific to each gene and treatment combination [33].

As gene expression studies become more popular, the complexity of the experiment increases. Instead of only simple treatment and control experiments, two or more factor experiments are being conducted. This increase in experiment complexity has led to many scientific questions involving the hypothesis testing of an interaction between two factors. For example, testing a probe by genotype interaction can result in inferences about polymorphism in the probe, such as single nucleotide polymorphism (SNP) and insertion-deletion (indel) [34-37]; testing a probe by sex can imply that alternative splicing occurs between male and female subjects [38]; and in pharmacogenomic studies, testing the genotype-drug/treatment or genotype-disease interaction may be of interest [39]. Thus far, all the development of ANOVA methods for microarray studies has focused on tests of main effects.

Here, a generalized shrinkage estimator incorporating both within- and between-gene heteroscedasticities is developed (see Lehmann and Cesella [40] for a review of shrinkage estimation). In any given experiment, both within-gene and between-gene heteroscedasticity may exist; thus, taking these possibilities into account should

lead to an improved test statistic. Moreover, given the increasing complexity of recent studies and the burgeoning interest in hypotheses that involve interactions, we focus on an improved shrinkage-based F -test for interaction terms.

Methods

Here we develop new shrinkage estimates for the error term and show how to use these estimates to construct F -like statistics. We then estimate the null distribution of these statistics by using permutation tests.

Shrinkage error estimators

Shrinkage error estimators pull individual error estimates toward shrinkage targets, with the amount of shrinkage depending on the variability of individual error estimates [28,40]. Let the gene-specific error estimates for all genes i and subgroups k be

$$(\hat{\sigma}_{1,1}^2, \dots, \hat{\sigma}_{1,K}^2, \dots, \hat{\sigma}_{I,K}^2),$$

$i = 1, \dots, I$, $k = 1, \dots, K$, and let $\sigma_{i,k}^2$ be the true variance of gene i in group k . When the experimental design is balanced, $\hat{\sigma}_{i,k}^2$ is the residual mean square for gene i in group k and $v\hat{\sigma}_{i,k}^2/\sigma_{i,k}^2 \sim \chi_v^2$, where v represents the degrees of freedom for the error estimates.

The choices of shrinkage targets in microarray data include the following:

1. Specific values for each gene-group combination
2. Gene-specific values that are the same across all other groups
3. Group-specific values that are the same across genes but different across groups
4. A single point representing the underlying common error

Correspondingly, these targets are correct when (1) there are both within-gene and between-gene heteroscedasticity; (2) there is only between-gene heteroscedasticity; (3) there is only within-gene heteroscedasticity; and (4) all error variances are identical. We now develop a generalized shrinkage error estimator using these four shrinkage targets.

Let $X_{i,k} \equiv \log \hat{\sigma}_{i,k}^2 - m \sim \log \sigma_{i,k}^2 + \log \chi_v^2/v - m$, where m is the mean of $\log \log \chi_v^2/v$. Then using asymptotic normal approximation of $X_{i,k}$, the distribution of $X_{i,k}$ s with different shrinkage targets for different gene i and group k combinations is

$$\begin{aligned} X_{i,k} | \theta_{i,k} &\sim N(\theta_{i,k}, \sigma^2) \\ \theta_{i,k} &\sim N(\mu + \alpha_i + \beta_k, \tau^2), \end{aligned} \quad (1)$$

where $\tilde{\theta} = (\theta_{1,1}, \dots, \theta_{1,K}, \dots, \theta_{I,1}, \dots, \theta_{I,K})$, $\tilde{\alpha} = (\alpha_1, \dots, \alpha_I)$ represents the gene-specific mean differences, and $\tilde{\beta} = (\beta_1, \dots, \beta_K)$ models different means with respect to different classes of the subgroups.

If σ^2 and τ^2 are known, then the Bayes estimator of $\theta_{i,k}$ under the squared error loss is [39]:

$$\theta_{i,k}^B = \frac{\sigma^2}{\sigma^2 + \tau^2} (\mu + \alpha_i + \beta_k) + \frac{\tau^2}{\sigma^2 + \tau^2} X_{i,k}.$$

Here, σ^2 is the variance of $\log \chi_v^2/v$ and is known [28,40], but τ^2 is not known. However, the marginal distribution of $X_{i,k}$ can be used to create an empirical Bayes estimator of τ^2 and hence of $\theta_{i,k}$. Marginally, $X_{i,k} \sim N(\mu + \alpha_i + \beta_k, \sigma^2 + \tau^2)$, $i = 1, \dots, I$, $k = 1, \dots, K$, and, from this model, the least square estimates of $\mu, \tilde{\alpha}, \tilde{\beta}, \hat{\mu}, \hat{\alpha}, \hat{\beta}$, are the uniformly minimum-variance and unbiased estimators. Using the fact that

$$E\left(\frac{[IK - (I + K - 1) - 2]}{\Sigma(X_{i,k} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_k)^2}\right) = \frac{1}{\sigma^2 + \tau^2},$$

the empirical Bayes estimator for τ^2 is $\Sigma(X_{i,k} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_k)^2 / [IK - (I + K - 1) - 2] - \sigma^2$.

Then, we can construct the positive-part empirical Bayes estimator [40]:

$$\theta_{i,k}^{EB+} = \hat{X}_{i,k} + \left(1 - \frac{[IK - (I + K - 1) - 2]\sigma^2}{\Sigma(X_{i,k} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_k)^2}\right)_+ \hat{X}_{i,k}$$

$$\hat{X}_{i,k} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_k,$$

where $(x)_+ = \max(x, 0)$. The generalized shrinkage error estimate for $\sigma_{i,k}$ can be obtained through exponentiating $\theta_{i,k}^{EB+}$ as follows:

$$\tilde{\sigma}_{Gen,i,k}^2 = \exp(\theta_{i,k}^{EB+}). \quad (2)$$

Using a similar argument, the generalized shrinkage error estimator with the shrinkage target at each gene is

$$\tilde{\sigma}_{Gen-gene,i,k}^2 = \exp(m + \hat{\mu} + \hat{\alpha}_i)$$

$$* \exp\left[\left(1 - \frac{[IK - (I - 1) - 2]\sigma^2}{\Sigma(X_{i,k} - \hat{\mu} - \hat{\alpha}_i)^2}\right)_+ (X_{i,k} - \hat{\mu} - \hat{\alpha}_i)\right], \quad (3)$$

with the shrinkage target at each group is

$$\tilde{\sigma}_{Gen-grp,i,k}^2 = \exp(m + \hat{\mu} + \hat{\beta}_k)$$

$$* \exp\left[\left(1 - \frac{[IK - (K - 1) - 2]\sigma^2}{\Sigma(X_{i,k} - \hat{\mu} - \hat{\beta}_k)^2}\right)_+ (X_{i,k} - \hat{\mu} - \hat{\beta}_k)\right], \quad (4)$$

and with the shrinkage target at the common error, we have

$$\tilde{\sigma}_{Gen-ce,i,k}^2 = \exp(m + \hat{\mu})$$

$$* \exp\left[\left(1 - \frac{[IK - 3]\sigma^2}{\Sigma(X_{i,k} - \hat{\mu})^2}\right)_+ (X_{i,k} - \hat{\mu})\right]. \quad (5)$$

The shrinkage error estimator proposed by Cui *et al.* [28] shrinks the gene-specific error estimators toward their common corrected geometric mean. Specifically, the estimator for σ_i^2 is calculated as

$$\tilde{\sigma}_{Cui,i}^2 = \exp\left(m + \frac{\Sigma X_i}{I}\right)$$

$$* \exp\left[\left(1 - \frac{[I - 3]\sigma^2}{\Sigma(X_i - \frac{\Sigma X_i}{I})^2}\right)_+ (X_i - \frac{\Sigma X_i}{I})\right], \quad (6)$$

where X_i is the residual variance estimate from a gene-specific model, and m and σ^2 are the mean and variance of $\log \frac{\chi^2 K v}{K v}$. The underlying assumption for

this estimator is that there is no between-gene heteroscedasticity, as this estimator shrinks every gene-specific error estimator toward one target. Therefore, it will overshrink the gene-specific error estimates when gene heteroscedasticity exists. In comparison, generalized shrinkage error estimators are flexible in terms of incorporating a different type of heteroscedasticity. Some degrees of freedom are used for incorporating the heteroscedasticity. However, the gain is that the error estimator is then closer to the underlying distribution and should lead to better performance of the resultant F-like test statistics as shown in the results section.

In formulas (2), (3), (5), and (6), m is the mean and σ^2 is the variance of a log-transformed chi-square random variable. The simulation-based approximate values of m and σ^2 can be found from Table 1 in work of Cui *et al.* [28]. Pounds [41] gave analytical expressions for these parameters and developed R code for the exact calculation. Here, the simulation-based approximate values were used.

Shrinkage F-like statistics

To construct a statistic for the hypothesis test of no interaction between two fixed effects, the traditional F -test is simply the ratio of the mean square of the interaction term (MSI) and the mean square of residuals (MSE). This F -test, referred to as F_1 [42], is $F_1 = \frac{MSI}{MSE} = \frac{MSI}{\hat{\sigma}^2}$. The F_1 test corresponding to a specific gene i is denoted by

$$F_{1,i} = \frac{MSI_i}{\hat{\sigma}_i^2}. \quad (7)$$

Table 1 Results from raw data permutation

Restricted?	Data set	F_1	F_2	F_3	F_{Cui}	F_{Gen}	$F_{Gen-gene}$	$F_{Gen-grp}$
YES	null-ce	5.05(0.07)	5.06(0.08)	5.12(0.17)	5.09(0.10)	5.09(0.10)	5.05(0.08)	5.11(0.10)
	null-gh	5.02(0.07)	5.13(0.16)	5.26(0.20)	5.03(0.07)	5.07(0.12)	5.03(0.07)	5.11(0.16)
	null-wgh	4.97(0.07)	4.96(0.09)	4.93(0.18)	4.99(0.08)	4.99(0.12)	4.96(0.09)	5.01(0.16)
	null-bgh	5.02(0.07)	4.99(0.17)	5.03(0.21)	5.02(0.07)	5.02(0.15)	5.01(0.09)	5.03(0.18)
NO	null-ce	5.10(0.07)	5.06(0.08)	5.06(0.08)	7.4(0.12)	5.15(0.09)	5.12(0.08)	5.08(0.08)
	null-gh	5.08(0.07)	5.12(0.16)	5.12(0.12)	7.4(0.09)	5.10(0.11)	5.07(0.10)	5.12(0.09)
	null-wgh	12.31(0.10)	7.56(0.10)	4.61(0.10)	17.37(0.14)	5.32(0.11)	5.07(0.09)	5.87(0.11)
	null-bgh	12.31(0.11)	6.63(0.17)	5.55(0.19)	15.68(0.12)	6.30(0.12)	6.10(0.11)	6.30(0.11)

CWER obtained from 1,000 permutations with the nominal significance level setting at 0.05, with standard errors in parentheses. Nine hundred simulation runs were performed to get empirical average CWER of all types of F -like test statistics.

The error variance estimator in this test uses data from only gene i . In oligonucleotide microarray models, the degrees of freedom for the error estimate can be small because the sample size of RNA is usually small, and hence the power of F_1 can be limited.

Following the method of constructing an F -test statistic given by Neter *et al.* [42], the gene-specific shrinkage F -like statistics for testing an interaction between two fixed effects can be obtained as

$$F_{Gen,i} = \frac{MSI_i}{\sum_k \tilde{\sigma}_{Gen,i,k}^2 / K},$$

$$F_{Gen-gene,i} = \frac{MSI_i}{\sum_k \tilde{\sigma}_{Gen-gene,i,k}^2 / K},$$

$$F_{Gen-grp,i} = \frac{MSI_i}{\sum_k \tilde{\sigma}_{Gen-grp,i,k}^2 / K},$$

$$F_{Gen-ce,i} = \frac{MSI_i}{\sum_k \tilde{\sigma}_{Gen-ce,i,k}^2 / K},$$

$$F_{Cui,i} = \frac{MSI_i}{\tilde{\sigma}_{Cui,i}^2}.$$

When the homoscedastic error assumption is true, the pooled variance estimator, $\hat{\sigma}_{pool}^2$, can be used to construct an F -like statistic. For a balanced design, the pooled variance estimate is the average of all gene-specific error estimates. This statistic is denoted by F_3 using the same notation used by Cui and Churchill [22], who also introduced another shrinkage-type F statistic, F_2 , which can also borrow information across genes when estimating the residual variances. The statistic F_2 uses an equal-weighted average of a gene-specific error estimator $\hat{\sigma}^2$ and $\hat{\sigma}_{pool}^2$. The definitions of $F_{2,i}$ and $F_{3,i}$ are

$$F_{2,i} = \frac{MSI_i}{0.5\hat{\sigma}_i^2 + 0.5\hat{\sigma}_{pool}^2},$$

$$F_{3,i} = \frac{MSI_i}{\hat{\sigma}_{pool}^2}.$$

Permutation tests

For the proposed generalized shrinkage F -like test statistics, the null distributions are not known named distributions. Therefore, an empirical approach such as a permutation test can be used to estimate the null distributions. The permutation test for interaction is complicated, because there is no exact permutation test for such a purpose [43]. We therefore must consider an approximate permutation method for testing an interaction term in a crossed fixed/mixed model [44,45].

Permutation approaches developed previously focused on a single ANOVA model. In the typical gene expression study, thousands of ANOVA models are considered simultaneously. The additional complexity of the shrinkage F -like statistics indicates that Monte Carlo studies are needed to investigate the performance of residual permutation and raw data permutation, with restrictions or not, in a gene-expression analysis. The choice of permutation procedures is critical for assessing the performance of a test statistic.

For all the modified F -like statistics presented in the previous section, the null distributions can only be approximated empirically, but permutation procedures can be used to find the approximate null distribution of all the F and F -like statistics. The important issues in performing a permutation analysis include the choice of the exchangeable units under the null hypothesis, the choice of using restricted permutation or not, and the choice of residual permutation or raw data permutation. These choices influence the power of a test statistic.

Residual permutation using residuals from a reduced model and unrestricted raw data permutation can be used to approximate the null distribution of a statistic for testing an interaction term [44]. When using F_1 to test an interaction term in a single ANOVA model, the residual permutation leads to a more powerful test than unrestricted raw data permutation [44]. However, in gene expression analysis, thousands of gene-specific ANOVA models are simultaneously considered, and for

a particular gene-specific ANOVA model, information from other gene-specific ANOVA models is used to construct the shrinkage error estimate. Hence, both residual permutation and raw data permutation were investigated. Furthermore, both restricted and unrestricted permutations were studied, because the permutation units are exchangeable only within each particular group when within-gene heteroscedasticity is present across those subgroups.

Results

The properties of this shrinkage estimator are compared with those of other existing F and F -like statistics that have been proposed and described in the “Shrinkage F -like statistics” section.

Simulation studies

The purpose of these simulation studies was to compare the performances of F_1 , F_2 , F_3 , F_{Cui} , F_{Gen} , $F_{Gen-gene}$, and $F_{Gen-grp}$ in terms of type I error and power and to compare the results of a particular F -like statistic using four different permutation strategies: restricted/unrestricted residual permutation and restricted/unrestricted raw data permutation.

In these simulation studies, 100 genes with two probes for each gene and three replicates from each of two lines were simulated to mimic a split-plot design in a general oligonucleotide microarray experiment. The gene-specific ANOVA model in which data were generated from the model, $y_{plr} = P_p + L_l + RL_{rl} + PL_{pl} + \epsilon_{plr}$, $wp = 1, 2$, $l = 1, 2$, $r = 1, 2, 3$, where P , L , RL , and PL represent probe, line, replicates from a particular line, and the interaction between probe and line, respectively.

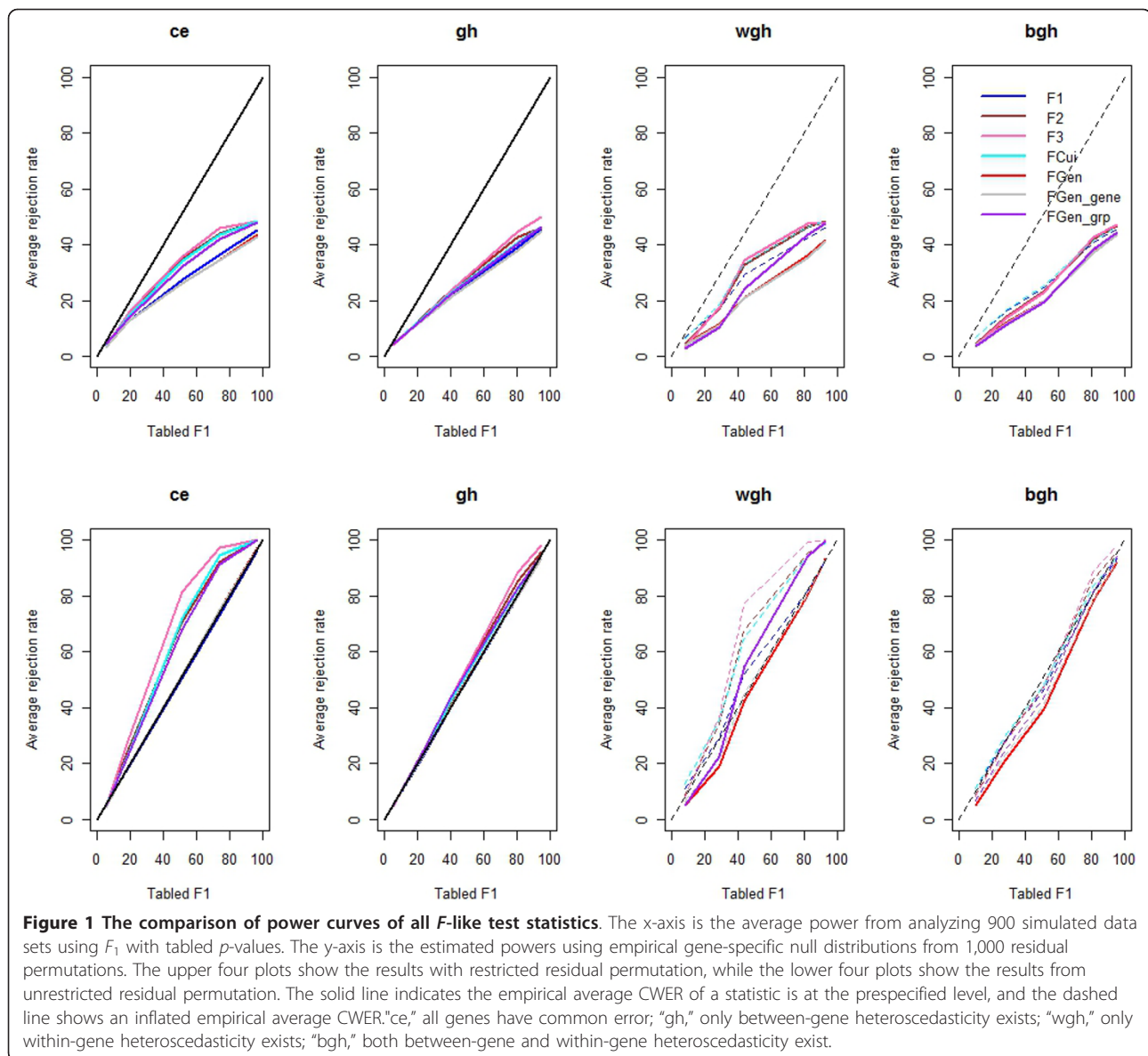
Replicates were nested within each line, and RL is usually treated as a random effect during the model-fitting procedure, which results in a correlation between probes from the same biological sample. In the simulated data sets, the correlation between genes was 0. As many as 900 simulation runs were carried out to compare the performances of F_1 , F_2 , F_3 , F_{Cui} , F_{Gen} , $F_{Gen-gene}$, and $F_{Gen-grp}$ based on different permutation procedures. The four permutations tested were unrestricted residual permutation, restricted residual permutation with respect to each line, unrestricted raw data permutation, and restricted raw data permutation with respect to each line. The residuals permuted were from a reduced fixed model with fixed effects for only line and probe.

Two types of data were simulated: null cases and cases with a probe by line interaction at a range of degrees. Null cases included: *null-ce*, all probe-level expression values were simulated from the standard normal distribution; *null-gh*, the gene-specific error variances were simulated from the log-normal distribution with mean log at 0 and standard deviation at 2, mimicking the

general heteroscedastic error distribution in typical datasets; *null-wgh*, all genes had the same error structures and the residual error variance of line 1 was 100 times that of line 2; *null-bgh*, simulated data were modified from *null-gh*, with the variance of line 1 multiplied by 100. Correspondingly, *ce*, *gh*, *wgh*, and *bgh* in Figures 1 and 2 were simulated by adding interaction terms to *null-ce*, *null-gh*, *null-wgh*, and *null-bgh*. Quantitative interaction was assumed and the differences in the opposite direction were set to make the detection powers for an interaction term based on traditional F -statistics and tabled p -values range from 0.05 to 0.95.

Tables 1 and 2 show the results from 900 simulation runs using raw data permutation and residual permutation, respectively. Data in Table 1 suggest that when both types of gene heteroscedasticity exist, the unrestricted raw data permutation had a greater average comparison-wise error rate (CWER) than residual permutation. Raw data permutation with restriction can control prespecified CWER in all cases. In Table 2, for the common error cases, all test statistics had the prespecified CWER from both restricted and unrestricted residual permutation. When within-gene heteroscedasticity existed, F_1 and F_{Cui} had inflated CWER from both two residual permutation tests. Restricted residual permutation reduces, but does not solve, this problem. For F_2 and F_3 , only the restricted residual permutation could control the prespecified CWER. For F_{Gen} , $F_{Gen-gene}$, and $F_{Gen-grp}$, restricted residual permutation gave conservative results in terms of having CWER smaller than the prespecified level. When the shrinkage target is correctly set, unrestricted residual permutation controls the nominal CWER. As expected, only F_{Gen} coupled with unrestricted residual permutation could be used for all cases, because the CWER was always less than the nominal level.

Further simulations to compare the rejection rates were conducted. Only results from residual permutation are shown because it was found that raw data permutation was less powerful than residual permutation. This is consistent with the findings of Anderson and Ter Braak [44]. Figure 1 shows the estimated average null hypothesis rejection rate curves from all F -like statistics and both restricted and unrestricted residual permutation procedures. The x-axis represents the average null hypothesis rejection rate using F_1 and the tabulated p -values. The solid line shows that the corresponding statistic controls the prespecified CWER, and the dashed line shows that the corresponding CWER was inflated. In general, restricted residual permutation is less powerful than unrestricted residual permutation. For example, the power of all statistics from unrestricted residual permutation almost doubled in some cases where heteroscedasticity existed.



When the common error assumption is valid, F_3 is obviously the most powerful test and the prespecified CWER is controlled. All other F -like statistics performed very similarly in this case. When the shrinkage target was correctly set, the resultant test statistic was the most powerful one. For example, when there was only within-gene heteroscedasticity, $F_{Gen-grp}$ was more powerful than F_{Gen} and $F_{Gen-gene}$ based on either restricted or unrestricted residual permutation. The rejection rate comparison of statistically valid test statistics is further illustrated in Figure 2, where the x-axis is the average rejection rate from using F_{Gen} and unrestricted residual permutation. Figure 2 clearly shows that unrestricted residual permutation is more favorable in terms of power. $F_{Gen-grp}$ appears to be more powerful than F_{Gen}

but when both types of gene heteroscedasticities occur, $F_{Gen-grp}$ has inflated CWER.

Drosophila data

The data used in this study are from a gene expression comparison study between *D. melanogaster* and *D. simulans* [46]. Expression of 10 genotypes of each species was measured in male flies. In *D. simulans*, each genotype was measured separately, and in *D. melanogaster*, a pool of 10 genotypes was measured. All genotypes (individual or pooled) were independently isolated and hybridized three times. The goal of the original study was to provide a genome-wide approach to identifying candidate genes potentially responsible for adaptation and speciation in *D. simulans* and *D. melanogaster*. In this study, we focus on

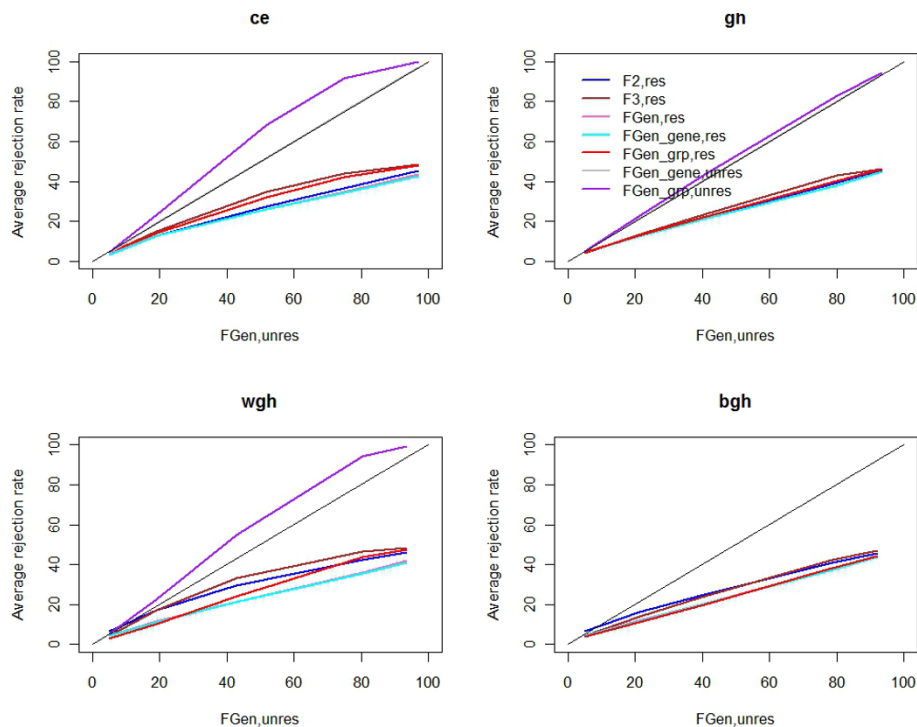


Figure 2 The comparison of power curves of F_{Gen} from unrestricted residual permutation versus other F -like test statistics. Only results from permutation combinations that can control prespecified CWER are used in this figure. The x-axis is the average power after analyzing 900 simulated data sets using F_{Gen} and 1,000 unrestricted residual permutations. The y-axis is the estimated power from other F -like test statistics and empirical gene-specific null distributions based on the appropriate permutation. The solid black line corresponds to F_{Gen} with unrestricted permutation, and this test always controls prespecified CWER. "ce," all genes have common error; "gh," only between-gene heteroscedasticity exists; "wgh," only within-gene heteroscedasticity exists; "bgh," both between-gene and within gene heteroscedasticity exist; "res," restricted permutation; "unres," unrestricted permutation.

identifying sequence differences between genotypes in *D. simulans* based on hybridization profiles. Within-gene heteroscedasticity is expected because the genotypes come from different lines. The proposed generalized shrinkage F -like test statistics F_{Gen} , $F_{Gen-gene}$, and $F_{Gen-grp}$ were compared with F_2 , F_3 with restricted residual permutation, which could control prespecified CWER for any variance structure in simulation studies. Furthermore, Smyth's moderated F -test statistic [25] without multiple testing

adjustment and controlling the false discovery rate (FDR) at 5% were used for comparison. As the main interest is in sequence difference, the focus is on the test of interaction between line and probe. The split plot model described above is used. SAS program codes are included in the additional files (additional file 1 and additional file 2).

The *Drosophila* genome has been fully sequenced and both SNPs and indels can cause a significant interaction term. Thus, the false positive rate and detection power

Table 2 Results from residual permutation

Restricted?	Data set	F_1	F_2	F_3	F_{Cui}	F_{Gen}	$F_{Gen-gene}$	$F_{Gen-grp}$
YES	null-ce	4.59(0.07)	4.15(0.08)	3.63(0.14)	4.09(0.09)	4.55(0.07)	3.23(0.06)	4.44(0.08)
	null-gh	4.57(0.07)	4.1(0.13)	3.95(0.16)	4.49(0.07)	4.6(0.07)	4.61(0.07)	4.38(0.07)
	null-wgh	6.74(0.08)	4.33(0.08)	3.51(0.14)	6.49(0.09)	4.38(0.07)	4.2(0.07)	2.78(0.09)
	null-bgh	6.74(0.08)	4.35(0.16)	4.07(0.19)	6.58(0.08)	4.36(0.07)	4.16(0.07)	3.64(0.07)
NO	null-ce	5.1(0.07)	4.99(0.08)	4.5(0.08)	4.59(0.08)	4.99(0.07)	4.1(0.07)	4.68(0.07)
	null-gh	5.1(0.07)	4.83(0.1)	4.59(0.11)	5.08(0.07)	4.99(0.07)	5.01(0.07)	4.95(0.07)
	null-wgh	10.75(0.09)	8.46(0.09)	7.6(0.09)	12.37(0.11)	5.03(0.08)	6.43(0.08)	4.93(0.08)
	null-bgh	10.75(0.1)	8.38(0.17)	8.07(0.19)	10.79(0.1)	5.02(0.08)	6.38(0.08)	6.73(0.08)

CWER obtained from 1,000 permutations with the nominal significance level setting at 0.05, with standard errors in parentheses. Nine hundred simulation runs were performed to get empirical average CWER of all types of F -like test statistics.

based on SNP/indel sequence information can be calculated for a subset of the data. In the data set, there were 10 lines from *D. simulans* and three replicates from each line. Each probe set had 14 probes. The 1,285 probesets containing all “good” probes were selected. A “bad” probe’s sequence satisfies one or more of the following criteria: it matches the *D. simulans* genome multiple times; it cannot be mapped to the flybase 4.2.1 genome; or, it has no information, such as hitting outside an exon, hitting a poorly aligned region, or hitting a region lacking a sequence. SNP or indel information could be determined in 777 probesets. For this data set, there was a high degree of within-gene heteroscedasticity: about 22.3% of the probe sets had a difference in line-specific residual variance estimates as large as or more than a 10-fold change. Therefore, as suggested by the conclusions from simulation studies, unrestricted residual permutation and restricted residual permutation were used for generalized shrinkage *F*-like test statistics (F_{Gen} , $F_{Gen-gene}$, $F_{Gen-grp}$) and restricted residual permutation was used for statistics (F_2 , F_3). The results are shown in Table 3. Consistent with the findings from the simulation studies, F_{Gen} had about 30% more detecting power by valuing the within-gene heteroscedasticity than the other *F*-like test statistics (F_2 , F_3). The false discovery rate of F_{Gen} was slightly higher than that of F_2 , F_3 . $F_{Gen-gene}$ and $F_{Gen-grp}$ performed similarly to F_{Gen} . Both of Smyth’s moderated *F*-test statistic without multiple testing adjustment and with FDR set at 5% for multiple testing adjustment detected more SNPs and indels but at the expense of a greater FDR than F_{Gen} .

Discussion

For gene expression analysis, ANOVA models have been a popular modeling technique. Based on ANOVA models, flexible shrinkage *F*-like test statistics were developed to account for both the within-gene and between-gene heteroscedasticities. The emphasis here is on

testing an interaction term, as this case is of increasing interest to biologists, and there is no clear existing theory on the most powerful, valid approach for such statistics. For all *F*-like statistics studied here, their null distributions were approximated empirically through permutations. Four different permutation procedures were investigated for eight different *F*-like statistical tests of the interaction term.

As expected, we found that when an error estimator overshrinks, the resulting *F*-like statistic cannot control the prespecified CWER. For example, $F_{Gen-gene}$ is an over-shrinkage error estimator when there is within-gene heteroscedasticity. As a result, compared with generalized shrinkage *F*-like statistics, it is not valid when within-gene heteroscedasticity exists. Undershrinkage is also important, as it will lead to a conservative test and lower power. This is clearly demonstrated when the common error can be assumed and the most powerful valid test is $F_{Gen-grp}$.

The most striking result was the impact of the permutation procedures. Although this was not completely unexpected [43-45], the effect of the permutation procedures is dramatic and worthy of special attention. Unrestricted raw data permutation could not control prespecified CWER when there was within-gene heteroscedasticity. Restricted raw data permutation could be used, but it was less powerful than residual permutation. Also consistent with findings from Anderson and Ter Braak [44], restricted permutations are less powerful than unrestricted permutations. However, unrestricted permutations are valid only for a common error and when between-gene heteroscedasticity exists for our proposed shrinkage statistics; they are not valid in combination with F_2 , F_3 , or F_{Cui} . For $F_{Gen-grp}$, the unrestricted permutation can also be used in cases having within-gene heteroscedasticity, while only F_{Gen} is valid with unrestricted permutation in all cases in terms of controlling prespecified CWER. Interestingly,

Table 3 Probe sets with significant *line*probe* terms found by *F*-like test statistics and appropriate residual permutation procedures and Smyth’s moderated *F*-test statistic

Test statistic	Restricted permutation?	Number of probe sets found	True false discovery rate	Power
F_2	Yes	124	22.6%	12.4%
F_3	Yes	187	29.4%	17.1%
F_{Gen}	No	453	29.5%	41.1%
$F_{Gen-gene}$	No	455	28.8%	41.7%
$F_{Gen-grp}$	No	474	28.9%	43.4%
F_{Gen}	Yes	136	24.3%	13.3%
$F_{Gen-gene}$	Yes	122	22.1%	12.2%
$F_{Gen-grp}$	Yes	116	21.5%	11.7%
moderatedF - 1	N/A	535	34.1%	75.5%
moderatedF - 2	N/A	813	34.4%	68.8%

The CWER was set to 0.05. Gene-specific cutoff values were obtained from 1,000 permutations. “moderated F-1” and “moderated F-2” represent results from using moderated *F* statistic without any multiple testing adjustment and setting FDR to 5%.

the power gain from using the correct shrinkage target $F_{Gen-grp}$ rather than F_{Gen} is far less than that of using unrestricted permutation. The result is that F_3 is never the most powerful choice when testing an interaction term.

The correct shrinkage target can lead to the most powerful test statistic. As one of the reviewers suggested, a statistical test may be applied to help pick the best shrinkage target before obtaining shrinkage error estimates. However, this extra testing step may inflate the CWER of the test statistic when there is gene heteroscedasticity. For example, when there are both types of gene heteroscedasticities, it is possible that the above test suggests only within-gene heteroscedasticities exist, and $F_{Gen-grp}$ is shown to inflate the CWER. There is minimal penalty to using the shrinkage estimator we propose, so we recommend setting the shrinkage target in the full space spanned by group and gene and using unrestricted permutation to compensate for the possible power loss in fewer degrees of freedom left for estimating the errors.

Conclusions

The proposed generalized shrinkage F -like statistic with shrinkage targets located in a space spanned by gene and another group, F_{Gen} , with unrestricted residual permutation is always valid in terms of having a prespecified CWER. This statistic has reasonable power in most cases; thus, it is generally recommended to be applied to test an interaction term in the analysis of real gene expression data.

Additional material

Additional file 1: SAS program code 1. SAS program code for analyzing the real data set using residual permutation without restriction.

Additional file 2: SAS program code 2. SAS program code for analyzing the real data set using residual permutation with restriction.

List of abbreviations

CWER: comparison-wise error rate; FDR: false discovery rate; indel: insertion and deletion; SNP: single nucleotide polymorphism;

Acknowledgements

We thank Brandon Walts for identifying true SNP positions; Angela J. McArthur and David R. Galloway for their help in scientific editing; associate editor and three reviewers for their constructive comments that much improved this manuscript. This research was supported by NIH 1R01GM077618 (McIntyre), NIH 1R01GM081704 (Casella).

Author details

¹Department of Preventive Medicine, Stony Brook University, Stony Brook, NY 11794, USA. ²Department of Statistics, University of Florida, Gainesville, FL 32611, USA. ³Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL 32611, USA. ⁴The Genetics Institute, University of Florida, Gainesville, FL 32611, USA.

Authors' contributions

All authors contributed to the design of the overall strategy. JY carried out all the analysis and drafted the manuscript. LMM and GC helped to draft and finalize the manuscript. All authors read and approved the final manuscript.

Received: 10 June 2010 Accepted: 1 November 2011

Published: 1 November 2011

References

1. Fodor SPA: **Massively parallel genomics.** *Science* 1997, **277**:393-395.
2. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW: **Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.** *Proceedings of National Academy Science* 1996, **93**:10614-10619.
3. Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR: **Ploidy regulation of gene expression in yeast.** *Science* 1999, **285**:251-254.
4. Tu BP, Kudlicki A, Rowicka M, McKnight SL: **Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes.** *Science* 2005, **310**:1152-1158.
5. White KP, Rifkin SA, Hurban P, Hogness DS: **Microarray analysis of Drosophila development during metamorphosis.** *Science* 1999, **286**:2179-2184.
6. Chabas D, Baranzini SE, Mitchell D, Bernard CCA, Rittling SR, Denhardt DT, Sobel RA, Lock C, Karpuz M, Pedotti R, Heller R, Oksenberg JR, Steinman L: **The influence of the proinflammatory cytokine, Osteopontin, on autoimmune demyelinating disease.** *Science* 2001, **294**:1731-1735.
7. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi MY, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M: **Large-scale copy number polymorphism in the human genome.** *Science* 2005, **305**:525-528.
8. Blekhan R, Marioni JC, Zumbo P, Stephens M, Gilad Y: **Sex-specific and lineage-specific alternative splicing in primates.** *Genome Research* 2010, **20**(2):180-189.
9. Butte A: **The use and analysis of microarray data.** *Nature Reviews* 2002, **1**:951-960.
10. Churchill GA: **Fundamentals of experimental design for cDNA microarrays.** *Nature Genetics* 2002, **32**:490-495.
11. Craig BA, Black MA, Doerge RW: **Gene expression data: The technology and statistical analysis.** *Journal of Agricultural, Biological, and Environmental Statistics* 2003, **8**(1):1-28.
12. Allison DA, Cui X, Page GP, Sabripour M: **Microarray data analysis: from disarray to consolidation and consensus.** *Nature Reviews Genetics* 2006, **7**:55-65.
13. Kerr MK, Martin M, Churchill GA: **Analysis of variance for gene expression microarray data.** *Journal of Computational Biology* 2000, **7**:819-837.
14. Kerr MK, Churchill GA: **Statistical design and the analysis of gene expression microarrays.** *Genetical Research* 2001, **77**:123-128.
15. Kerr MK, Churchill GA: **Experimental design for gene expression microarrays.** *Biostatistics* 2001, **2**:183-201.
16. Pritchard CC, Hsu L, Delrow J, Nelson PS: **Project normal: Defining normal variation in mouse gene expression.** *Proceedings of the National Academy of Sciences USA* 2001, **98**:13266-13271.
17. Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Ashfari C, Paules RS: **Assessing gene significance from cDNA microarray expression data via mixed models.** *Journal of Computational Biology* 2001, **8**(6):625-637.
18. Kerr MK, Afshari CA, Bennett L, Bushel P, Martinez J, Walker NJ, Churchill GA: **Statistical analysis of a gene expression microarray experiment with replication.** *Statistica Sinica* 2002, **12**:203-217.
19. Wu H, Kerr MK, Cui XQ, Churchill GA: **MAANOVA: A Software package for the analysis of spotted cDNA microarray experiments.** In *The analysis of gene expression data: methods and software* Springer; 2002, 313-341.
20. Chu T, Weir B, Wolfinger R: **A systematic statistical linear modeling approach to oligonucleotide array experiments.** *Mathematical Biosciences* 2002, **176**:35-51.
21. Wayne ML, Pan YJ, Nuzhdin SV, McIntyre LM: **Additivity and transacting effects on gene expression in male Drosophila simulans.** *Genetics* 2004, **168**:1413-1420.
22. Cui X, Churchill GA: **Statistical tests for differential expression in cDNA microarray experiments.** *Genome Biology* 2003, **4**(4):201.

23. Baldi P, Long AD: **A Bayesian framework for the analysis of microarray expression data: Regularized t-test and statistical inferences of gene changes.** *Bioinformatics* 2001, **17**:509-519.
24. Lönnstedt I, Speed T: **Replicated microarray data.** *Statistica Sinica* 2002, **12**:31-46.
25. Smyth GK: **Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.** *Statistical Applications in Genetics and Molecular Biology* 2004, **3**, No. 1, Article 3.
26. Tong TJ, Wang YD: **Optimal shrinkage estimation of variances with applications to microarray data analysis.** *Journal of the American Statistical Association* 2007, **102**:113-122.
27. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to ionizing radiation response.** *The Proceedings of National Academy Science* 2001, **98**:5116-5121.
28. Cui X, Hwang JTG, Qiu J, Blades NJ, Churchill GA: **Improved statistical tests for differential gene expression by shrinking variance components estimates.** *Biostatistics* 2005, **6**:59-75.
29. Feng S, Wolfinger RD, Chu TM, Gibson GC, McGraw LA: **Empirical Bayes analysis of variance component models for microarray data.** *Journal of Agricultural, Biological & Environmental Statistics* 2006, **11**:13:197-190.
30. Kim SY, Lee JW, Sohn IS: **Comparison of various statistical methods for identifying differential gene expression in replicated microarray data.** *Statistical Methods in Medical Research* 2006, **15**:3-20.
31. Hwang JTG, Liu P: **Optimal tests shrinking both means and variances applicable to microarray data analysis.** preprint 2007-03 Department of Statistics, Iowa State University, IA; 2007.
32. Kizilkaya K, Tempelman RJ: **A general approach to mixed effects modeling of residual variances in generalized linear mixed models.** *Genetics Selection Evolution* 2005, **37**:31-56.
33. Jaffrezic F, Marot G, Degrelle S, Isabelle H, Foulley JL: **A structural mixed model for variances in differential gene expression studies.** *Genetical Research* 2007, **89**(1):19-25.
34. Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R: **Single-feature polymorphism discovery in the Barley transcriptome.** *Genome Biology* 2005, **6**:R54.
35. Kirst M, Caldo R, Casati P, Tanimoto G, Walbot V, Wise RP, Buckler ES: **Genetic iversity contribution to errors in short oligonucleotide microarray analysis.** *Plant Biotechnology Journal* 2006, **4**:489-498.
36. Zhang X, Shiu SH, Cal A, Borevitz JO: **Global analysis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis Thaliana Using whole genome tiling arrays.** *PLoS Genetics* 2008, **4**(3):e1000032.
37. Zhang X, Borevitz JO: **Global Analysis of Allele-specific Expression in Arabidopsis Thaliana.** *Genetics* 2009, **182**(4):943-954.
38. McIntyre LM, Bono LM, Genissel A, Westerman R, Junk D, Telonis-Scott M, Harshman L, Wayne ML, Kopp A, Nuzhdin SV: **Sex specific expression of alternative transcripts in Drosophila.** *Genome Biology* 2006, **7**:R79.
39. Kelly P, Zhou YH, Whitehead J, Stallard N, Bowman C: **Sequentially testing for a gene-drug interaction in a genomewide analysis.** *Statistics in Medicine* 2008, **27**:2022-2034.
40. Lehmann EL, Casella G: *Theory of Point Estimation*. 2 edition. New York: Springer-Verlag; 1998.
41. Pounds S: **Computational enhancement of a shrinkage-based analysis of variance F-test proposed for differential gene expression analysis.** *Biostatistics* 2007, **8**:3:505-506.
42. Neter J, Wasserman W, Kutner MH: *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*. 3 edition. Irwin, Inc; 1990.
43. Edgington ES: *Randomization Tests*. 3 edition. Marcel Dekker, New York; 1995, (1995).
44. Anderson MJ, Ter Braak CJF: **Permutation tests for multi-factorial analysis of anova.** *Journal of Statistical Computation and Simulation* 2003, **73**:2:85-113.
45. Churchill GA, Doerge RW: **Naive application of permutation testing leads to inflated type I error rates.** *Genetics* 2008, **178**:609-610.
46. Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM: **Common pattern of evolution of gene expression level and protein sequence in drosophila.** *Molecular Biology and Evolution* 2004, **21**:1308-1317.

doi:10.1186/1471-2105-12-427

Cite this article as: Yang et al.: Generalized shrinkage F-like statistics for testing an interaction term in gene expression analysis in the presence of heteroscedasticity. *BMC Bioinformatics* 2011 **12**:427.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

