Assessing Statistical Significance in Microarray Experiments Using the Distance Between Microarrays

Douglas Hayden[1]*, **Peter Lazar**[1], **David Schoenfeld**[1,2] for The Inflammation and the Host Response to Injury Investigators

1 Biostatistics Center, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 2 Harvard School of Public Health, Boston, United States of America

Abstract

We propose permutation tests based on the pairwise distances between microarrays to compare location, variability, or equivalence of gene expression between two populations. For these tests the entire microarray or some pre-specified subset of genes is the unit of analysis. The pairwise distances only have to be computed once so the procedure is not computationally intensive despite the high dimensionality of the data. An R software package, permtest, implementing the method is freely available from the Comprehensive R Archive Network at http://cran.r-project.org.

Introduction

We propose permutation tests based on the matrix of pairwise distances between microarrays to compare location, variability, or equivalence of gene expression between two populations. These tests can be applied to the entire genome or any subset of genes of interest. Thus they can be used as global tests of difference in gene expression or as a testing method applicable to gene set analysis.

These tests have several advantages over permutation tests based directly on the gene expression data. First, they are not computationally intensive because they reduce the high dimensional expression data to the low dimensional distance matrix and this only has to be done once. Second, this same reduction in dimensionality results in a reduction in the dimensionality of the potential nuisance parameters. Thus, the assumption of exchangeability between groups, which ensures the validity of the permutation test, only has to apply to the pairwise distances not to the entire microarray.

Many gene expression studies are designed to detect differential gene expression in two clinical or biological populations. Interest may focus on individual genes, specified groups of genes, or all genes sampled by a particular microarray. Early efforts were devoted to individual gene analysis [1]. A variety of test statistics were proposed and an eventual consensus emerged on the use of false discovery rate estimation to correctly account for multiple testing [2].

In order to incorporate existing biological information into the analysis of differential gene expression more recent work has been devoted to methods that treat sets of related genes as the unit of analysis [3]. Many methods and tools have been developed to analyze sets of genes [4,1,5]. These include methods which base inference on lists of genes which individually exceed a specified cut off threshold for significance and those which base inference on scores which combine information over the entire gene set [5]. Further, gene set methods can be divided into competitive and self-contained tests [3]. Competitive tests compare the specified gene set to the remaining genes outside the set while self-contained tests depend only on the specified genes. Competitive tests have been shown to be conceptually flawed [3] and will not be considered further.

A number of self-contained, multivariate, tests have been proposed [6,7,8]. Any multivariate test requires a test statistic which aggregates information over all genes in the gene set and a method of inference to determine if the observed magnitude of the test statistic is extreme under the null hypothesis of no between group difference in gene expression. The primary challenge in developing these multivariate tests is posed by having relatively few samples and many genes. In particular, asymptotic methods may not be applicable and the sample gene covariance matrix may not be full rank and is therefore not invertible.

Kong *et al.* [7] discuss several modifications to Hotelling's T^2 [9] and suggest reducing the gene set to the lower dimensional subspace of principal components in which the sample gene covariance matrix becomes invertible. A permutation test is used for inference.

Goeman *et al.* [6] circumvent the problem of inverting the sample gene covariance matrix by developing a test statistic derived from a score test for a logistic regression model. In this model the gene expression values are considered to be fixed constants and the group indicator variable is considered to be stochastic. Their statistic is proportional to Hotelling's T^2 under the assumption of equal variance and zero covariance across all genes. They present the statistic's asymptotic distribution but suggest a permutation test for inference in small samples.

Hummel *et al.* [8] propose a test statistic which combines residual sums of squares from gene-wise linear regression models, which may include additional covariates of interest, fit with and without the group effect. They present the asymptotic distribution of the test statistic under the assumption that the genes follow a multivariate normal distribution and use a shrinkage estimator to regularize the sample gene covariance matrix. They also propose a permutation test based on the residual gene expression derived by regressing out an intercept and the additional covariates gene by gene. In the absence of additional covariates, the resulting statistic can be shown to be permutationally equivalent to Hotelling's T^2 under the assumption of equal variance and zero covariance across all genes. Note that two statistics are permutationally equivalent [10] if they have the same (or reverse) ordering over all permutations and therefore always give the same permutation p-value.

In this paper we propose methods to test for a significant difference in gene expression between two study groups based on the matrix of pairwise distances between microarrays. In the following sections we present the test statistics given the assumption of Euclidean distance, additive errors, and a completely randomized design. We also present an application to respiratory recovery in trauma patients and extensions to paired and blocked designs as well as the use of a distance measure based on correlation.

Methods

Introduction

Consider an experiment comparing gene expression in two populations. The gene expression values, or some function of the gene expression values which measures the biological signal for each gene, can be represented by two groups of column vectors. We will represent these column vectors by $X_{1,1}, \ldots, X_{1,N_1}$ and $X_{2,1}, \ldots, X_{2,N_2}$, for groups 1 and 2 respectively where N_1 and N_2 are the number of arrays in groups 1 and 2 respectively. For simplicity we will refer to these vectors as microarrays. Let $D[X_{i,j}, X_{k,l}]$ be the dissimilarity or distance between two microarrays.

Inference concerning the location and variability of groups 1 and 2 can be based on the three means:

$$\bar{D}_{11} = \frac{2}{N_1(N_1 - 1)} \sum_{i < j \leq N_1} D[X_{1,i}, X_{1,j}] \qquad (1)$$

$$\bar{D}_{22} = \frac{2}{N_2(N_2 - 1)} \sum_{i < j \leq N_2} D[X_{2,i}, X_{2,j}] \qquad (2)$$

$$\bar{D}_{12} = \frac{1}{N_1 N_2} \sum_{i \leq N_1, j \leq N_2} D[X_{1,i}, X_{2,j}] \qquad (3)$$

Where

1. $\bar{D}_{11}$ is the mean distance between microarrays within group 1.
2. $\bar{D}_{22}$ is the mean distance between microarrays within group 2.
3. $\bar{D}_{12}$ is the mean distance between microarrays between groups 1 and 2.

Additive Errors and Euclidean Distance

In this section we derive the expected value of $\bar{D}_{11}$, $\bar{D}_{22}$, and $\bar{D}_{12}$ assuming a completely randomized design, an additive error model, and squared Euclidean distance. Under these assumptions we have:

$$X_{1,i} = \mu_1 + \varepsilon_{1,i} \qquad (4)$$

$$X_{2,i} = \mu_2 + \varepsilon_{2,i} \qquad (5)$$

where μ_1 and μ_2 are the respective mean vectors of the microarrays in groups 1 and 2 and $\varepsilon_{1,i}$ and $\varepsilon_{2,i}$ are random error vectors with expected value 0 and variance covariance matrices Σ_1 and Σ_2 respectively. The errors are assumed to be independent across microarrays.

The squared Euclidean distance between any two microarrays, $X_{i,j}$ and $X_{k,l}$ is

$$\left| X_{i,j} - X_{k,l} \right|^2 = \left(X_{i,j} - X_{k,l} \right)^T \left(X_{i,j} - X_{k,l} \right) \qquad (6)$$

and a simple calculation yields

$$E[\bar{D}_{11}] = 2 Tr[\Sigma_1] \qquad (7)$$

$$E[\bar{D}_{22}] = 2 Tr[\Sigma_2] \qquad (8)$$

$$E[\bar{D}_{12}] = |\mu_1 - \mu_2|^2 + Tr[\Sigma_1] + Tr[\Sigma_2] \qquad (9)$$

where $Tr[]$ is the trace operator which gives the sum of the diagonal elements of a matrix.

We can now define a test statistic to compare the location or variability of groups 1 and 2.

To compare location let:

$$\Delta_l = \bar{D}_{12} - \frac{\bar{D}_{11} + \bar{D}_{22}}{2} \qquad (10)$$

To compare variability let:

$$\Delta_v = \bar{D}_{11} - \bar{D}_{22} \qquad (11)$$

These test statistics have expected values:

$$E[\Delta_l] = |\mu_1 - \mu_2|^2 \qquad (12)$$

and:

$$E[\Delta_v] = 2(Tr[\Sigma_1] - Tr[\Sigma_2]) \qquad (13)$$

Inference concerning the magnitude of Δ_v and Δ_l can be made using a permutation test. Each permutation consists of assigning N_1 microarrays to group 1 and the remaining N_2 to group 2. Note

that for each permutation the pairwise distances are simply re-indexed, they do not have to be recalculated. Only the values of $\bar{D}_{11}$, $\bar{D}_{22}$, and $\bar{D}_{12}$ and Δ_v and Δ_l have to be recalculated based on the re-indexing.

Let Δ_v^{obs} and Δ_l^{obs} be the observed values of Δ_v and Δ_l and let Δ_v^* and Δ_l^* be the values from a permutation. If there are a total of B permutations, and assuming $\Delta_v^{obs} > 0$, then

$$p_v = \frac{Number\left[\Delta_v^* \geq \Delta_v^{obs}\right]}{B} \qquad (14)$$

is a one-sided p-value [11] for rejecting the null hypothesis that $Tr[\Sigma_1] = Tr[\Sigma_2]$. If $\Delta_v^{obs} < 0$ then the inequality in Equation 14 is simply reversed. Similarly

$$p_l = \frac{Number\left[\Delta_l^* \geq \Delta_l^{obs}\right]}{B} \qquad (15)$$

is a one-sided p-value for rejecting the null hypothesis that $\mu_1 = \mu_2$.

Sometimes investigators design an experiment to compare a new technical method, such as sample preparation or target hybridization, to a proven "gold standard". In such an experiment interest centers on showing that the new method is equivalent to the gold standard. To be equivalent it should not differ in mean and not exhibit greater variability. Assuming that the microarrays in group 1 were prepared using the gold standard, a summary statistic which can be used to reject the null hypothesis of equivalence is given by:

$$\Delta_e = \bar{D}_{12} - \bar{D}_{11} \qquad (16)$$

If this statistic is large, then group 2 either has a different mean or more variability than group 1. This can easily be seen from its expected value under mean squared Euclidean distance:

$$E[\Delta_e] = |\mu_1 - \mu_2|^2 + (Tr[\Sigma_2] - Tr[\Sigma_1]) \qquad (17)$$

Inference concerning the magnitude of Δ_e can also be made using a permutation test. It should be noted that the statistics Δ_v, Δ_l, and Δ_e are all special cases of Mantel's U statistic [12], for which he derives the permutation variance, and Δ_l is similar to a special case of the MRPP statistic [13].

Paired Design

In this section we consider modifications to the test statistics and permutation test to account for pairing in the experimental design. We still assume an additive error model and squared Euclidean distance.

Suppose the experimental design is paired so that each microarray in group 1 is paired with a microarray in group 2. In this case $N_1 = N_2 = N$ and we assume that

$$X_{1,i} = \mu_1 + B_i + \varepsilon_{1,i} \qquad (18)$$

$$X_{2,i} = \mu_2 + B_i + \varepsilon_{2,i} \qquad (19)$$

where B_i is a random vector, independent of all other quantities, with expected value 0 and variance covariance matrix Σ_B. The quantities μ_1, μ_2, $\varepsilon_{1,i}$, and $\varepsilon_{2,i}$ are again defined as in Section.

To account for the pairing we drop the within pair distance in our test statistics since it will be systematically less than the between pair distances. Thus we re-define $\bar{D}_{12}$ as:

$$\bar{D}_{12} = \frac{1}{N^2} \sum_{i \neq j} D\left[X_{1,i}, X_{2,j}\right] \qquad (20)$$

then:

$$E[\bar{D}_{11}] = 2Tr[\Sigma_1] + 2Tr[\Sigma_B] \qquad (21)$$

$$E[\bar{D}_{22}] = 2Tr[\Sigma_2] + 2Tr[\Sigma_B] \qquad (22)$$

$$E[\bar{D}_{12}] = |\mu_1 - \mu_2|^2 + Tr[\Sigma_1] + Tr[\Sigma_2] + 2Tr[\Sigma_B] \qquad (23)$$

so that Δ_v and Δ_l still have the same expected values given in Equations 13 and 12.

For this paired design, only paired microarrays are exchangeable and the permutation test must respect this structure. Thus, a permutation consists of exchanging microarrays across groups 1 and 2 within each pair of an arbitrary subset of paired microarrays.

Blocked Design

In this section we consider modifications to the test statistics and permutation test to account for blocking in the experimental design. We still assume an additive error model and squared Euclidean distance.

Suppose the experiment consists of K blocks of related microarrays. In this case $N_{1,k}$ and $N_{2,k}$ are the number of microarrays in block k in group 1 and 2 respectively. We also assume that for blocks $k = 1, \ldots, K$

$$X_{1,i,k} = \mu_1 + B_k + \varepsilon_{1,i} \qquad (24)$$

$$X_{2,i,k} = \mu_2 + B_k + \varepsilon_{2,i} \qquad (25)$$

where B_k is a random vector , independent of all other quantities, with expected value 0 and variance covariance matrix Σ_B. The quantities μ_1, μ_1, $\varepsilon_{1,i}$, and $\varepsilon_{2,i}$ are again defined as in Section.

For this blocked design we need to modify the mean distances $\bar{D}_{11}$, $\bar{D}_{22}$, and $\bar{D}_{12}$ so that they depend only on within block distances. That is:

$$\bar{D}_{11} = \frac{1}{N_{11}} \sum_{k=1}^{K} \sum_{i < j \leq N_{1,k}} D\left[X_{1,i,k}, X_{1,j,k}\right] \qquad (26)$$

$$\bar{D}_{22} = \frac{1}{N_{22}} \sum_{k=1}^{K} \sum_{i < j \leq N_{2,k}} D\left[X_{2,i,k}, X_{2,j,k}\right] \qquad (27)$$

$$\bar{D}_{12} = \frac{1}{N_{12}} \sum_{k=1}^{K} \sum_{i \leq N_{1,k}, j \leq N_{2,k}} D\left[X_{1,i,k}, X_{2,j,k}\right] \qquad (28)$$

where:

$$N_{11} = \frac{1}{2}\sum_{k=1}^{K} N_{1,k}(N_{1,k}-1) \qquad (29)$$

$$N_{22} = \frac{1}{2}\sum_{k=1}^{K} N_{2,k}(N_{2,k}-1) \qquad (30)$$

$$N_{12} = \sum_{k=1}^{K} N_{1,k}N_{2,k} \qquad (31)$$

The permutation test for this design must respect the blocked structure. Microarrays are exchanged across groups only within each block.

Negative Log Correlation Distance

There are various possible measures of distance between microarrays besides squared Euclidean distance. In this section we consider a measure based on the Pearson product moment correlation [14] which we will denote by $cor[\cdot]$. Assuming an additive error model and a completely randomized design, the correlation between two microarrays, $Z_1 = \mu_1 + \varepsilon_1$ and $Z_2 = \mu_2 + \varepsilon_2$, is

$$cor[Z_1,Z_2] = \frac{cov[\mu_1,\mu_2]+cov[\mu_1,\varepsilon_2]+cov[\mu_2,\varepsilon_1]+cov[\varepsilon_1,\varepsilon_2]}{\sqrt{\sigma_{\mu_1}^2+2cov[\mu_1,\varepsilon_1]+\hat{\sigma}_{\varepsilon_1}^2}\sqrt{\sigma_{\mu_2}^2+2cov[\mu_2,\varepsilon_2]+\hat{\sigma}_{\varepsilon_2}^2}} \qquad (32)$$

where $cov[\cdot]$ is the sample covariance, σ^2 is the variance, and $\hat{\sigma}^2$ is the sample variance. By dividing the numerator and denominator in Equation 32 by $\sigma_{\mu_1}\sigma_{\mu_2}$ and rearranging some terms we get:

$$cor[Z_1,Z_2] = (cor[\mu_1,\mu_2]+\delta)\sqrt{\hat{R}_1}\sqrt{\hat{R}_2} \qquad (33)$$

where $\hat{R}_i$ is the sample estimate of the reliability [15]

$$R_i = \frac{\sigma_{\mu_i}^2}{\sigma_{\mu_i}^2+\sigma_{\varepsilon_i}^2} \qquad (34)$$

and

$$\delta = \frac{cov[\mu_1,\varepsilon_2]+cov[\mu_2,\varepsilon_1]+cov[\varepsilon_1,\varepsilon_2]}{\sigma_{\mu_1}\sigma_{\mu_2}} \qquad (35)$$

with $E[\delta] = 0$.

If the correlation is positive and under the assumption that $E[\delta^2]$ is negligibly small then, if we define the distance between Z_1 and Z_2 by

$$D[Z_1,Z_2] = -logcor[Z_1,Z_2] \qquad (36)$$

then, the expected value of $D[Z_1, Z_2]$ is approximated by

$$-logcor[\mu_1,\mu_2] - \frac{E[log\hat{R}_1]+E[log\hat{R}_2]}{2} \qquad (37)$$

Defining Δ_v and Δ_l as above in Equations 11 and 10, these test statistics have expected values

$$E[\Delta_v] = E\left[log\frac{\hat{R}_2}{\hat{R}_1}\right] \qquad (38)$$

and

$$E[\Delta_l] = -logcor[\mu_1,\mu_2] \qquad (39)$$

so that Δ_v estimates the log of the between group reliability ratio and Δ_l estimates the negative log of the correlation between the group means.

Using the negative log correlation distance the three designs considered above need no modification except in the interpretation of the statistics Δ_v and Δ_l as presented in Equations 38 and 39. In the paired design it is convenient to think of the random pair effect, B_i, as being absorbed in the error term. In the blocked design it is convenient to think of the random block effect, B_k, as being absorbed in the mean.

Note that the reliability given by Equation 34, which can be interpreted as a signal to noise ratio, is a reasonable measure of the within group distance or variability, but it depends on both the mean, μ, and the error, ε. Thus a significant difference in the within group distance as measured by Δ_v can be due to a difference in either the mean or the error.

A commonly used distance measure is one minus the correlation between microarrays. This distance measure is nearly equivalent to the negative log correlation distance for values of the correlation near one. However, by using this distance, the corrected distance between groups, Δ_l, does not have the correct expected value and does not estimate the error free distance between the groups.

Permutation Test Assumptions

Permutation tests may place weaker distributional assumptions on the data than do parametric tests but they are not totally free of assumptions. In general they only test the global null hypothesis, $H_0 : F_1 = F_2$ where F_1 is the data generating distribution for group 1 and F_2 is the data generating distribution for group 2. In particular they cannot test that a specific parameter of the data generating distribution differs between two groups without highly restrictive assumptions about the remaining (nuisance) parameters.

They do have the desirable property of being exact tests provided that, under the null hypothesis, the observations are exchangeable so that the joint distribution of the combined data set is invariant under permutations of the observation labels. The permutation tests we propose first reduce the extremely high dimensionality of the data (and attendant high dimensionality of the nuisance parameters) to the low dimensional matrix of pairwise distances. Since inference is based solely on the distribution of these distances the global null hypothesis becomes the much less restrictive assumption that the pairwise distances are identically distributed. Thus, only pairwise distances need be exchangeable by permutation of observation labels.

In particular, under Euclidean distance and the additive error model proposed above, the location permutation test, which tests the null hypothesis, $H_0 : \mu_1 = \mu_2$, is exact under the relatively weak
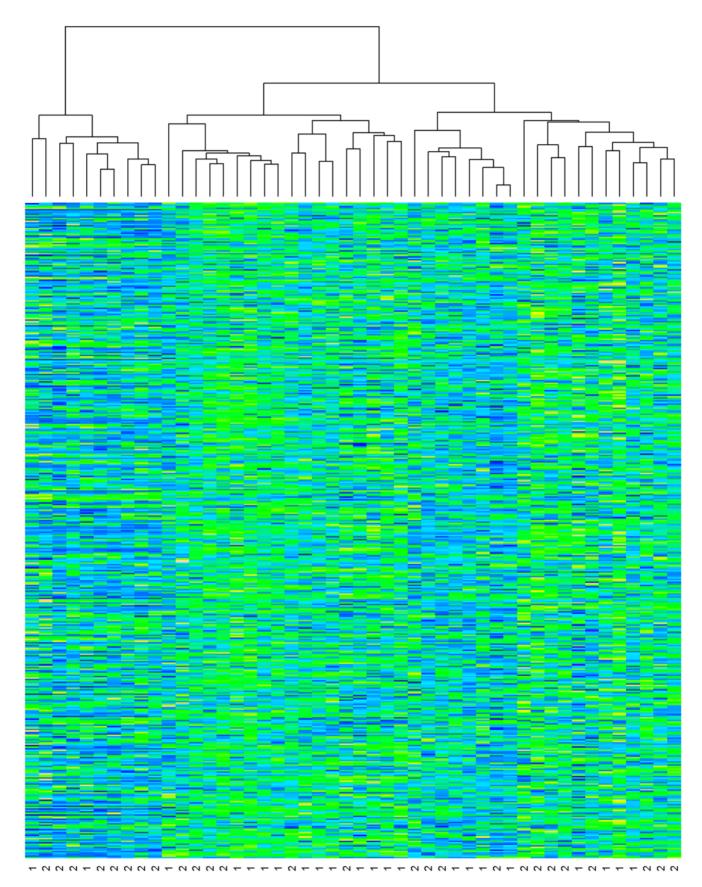
Figure 1. Heat Map
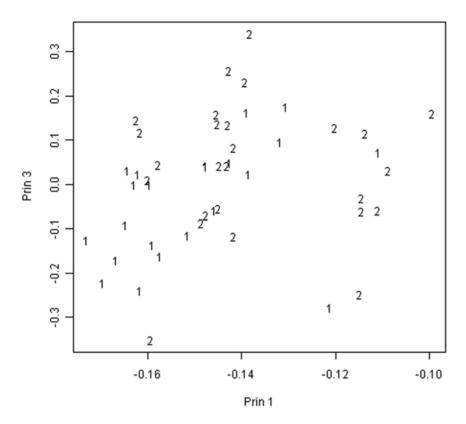doi:10.1371/journal.pone.0005838.g001

5

Figure 2. Principal Components
doi:10.1371/journal.pone.0005838.g002

assumption that the inner product of any two distinct error terms is exchangeable. That is, the test is exact if terms of the form $(\varepsilon_{i,j})^T(\varepsilon_{k,l})$, with $\{i,j\} \neq \{k,l\}$, are exchangeable across observations. These scalar terms are easily shown to be simply random errors in the sense that they each have expected value zero and are uncorrelated with all other terms.

Results

As an application of our proposed method we analyzed time to respiratory recovery in ventilated trauma patients in a data set previously described by Rajicic [16]. Patients were followed for 28 days post trauma and Affymetrix U133+2 microarrays were prepared from whole white blood cells sampled at days 0, 1, 4, 7, 14, 21, 28. We considered a subset of 48 ventilated patients who had a day one sample and divided them into two subgroups: those who recovered from ventilation prior to day seven (early recovery, N = 22) and those who did not (late recovery, N = 26). Of clinical interest is the potential association of inflammation on day one and subsequent respiratory recovery. To address this issue, a set of 445 probesets whose GO annotation included the term "inflammatory" was retrieved for analysis by a keyword search of the Affymetrix web site (http://www.affymetrix.com/index.affx).

Figure 1 shows a heat map of day one gene expression for the 48 patients (columns) over the 445 probesets (rows). In the figure, columns labeled with a "1" comprise the early recovery group, those labeled with a "2" comprise the late recovery group. Hierarchical clustering using Euclidean distance has been applied to the patients but not the probesets.

As can be seen in the figure, it is not readily apparent that the clustering separated the early and late recovery groups. However,

the highest split separated the patients into one group (right hand side of figure) having 20/38 (53%) early recovery patients while the remaining group (left hand side off figure) had only 2/10 (20%) early recovery patients. This imbalance may be suggestive of a difference in gene expression between the early and late recovery groups.

To formally test for a group effect we applied our proposed permutation test for a Euclidean distance location difference between the early and late recovery groups and obtained a one sided p-value of 0.0168, indicating a significant difference in gene expression.

As a check on this result, in the spirit of Kong's use of Hotelling's T-square in the principal component space [7], we applied MANOVA to the first three principal components. We obtained a two sided p-value of 0.0393 for the group effect, a result similar to the permutation test result. The choice of three principal components was arbitrary, however, and results vary with the number chosen for analysis.

Figure 2 illustrates the location of the two recovery groups in the subspace spanned by the first and third principal components. Note that the same cluster of two early recovery patients (denoted by "1") and eight late recovery patients (denoted by "2") can be seen on the right hand side of the figure.

Similar results were obtained using the globaltest R-package of Goeman et al. [6], with a p-value of 0.0167 as well as the GlobalAncova R-package of Hummel et al. [8], with a p-value of 0.0210.

Finally it may be of interest to note that, for the remaining 54230 probesets not associated with inflammation, the group difference was marginally significant with a p-value of 0.0462 by our method and p-values of 0.0508 and 0.0480 by the methods of

Goeman and Hummel respectively. This suggests that factors other than inflammation may also be associated with time to ventilator recovery.

Discussion

Our proposed test statistics attempt to divide differences in gene expression into differences in location and differences in variation. The hope is to find tests for one parameter when the other parameter differs. This is a common statistical problem that only has a solution under limited circumstances. For instance, if data are normally distributed it is possible to find an exact test for variability differences when there are differences in location but testing for differences in location when there are differences in variability is the famous Behrens-Fisher problem which has no optimal small sample solution [17]. We use permutation tests because the distribution functions of gene expression are unknown. The permutation tests we propose have the advantage that the pairwise distances are calculated before the permutations are applied so the tests are not computer intensive. The location test is designed to have power for a location difference and correct Type I error rate even given a variability difference, and the variability test is designed to have the opposite characteristics. Simulations (results not shown) show that, in the cases that we simulated, we succeeded. However, there may be circumstances where variability differences appear as location differences and vice-versa.

Since the distribution functions of gene expression are unknown it is not possible to calculate the power of the location test under specified alternative hypotheses. However, as pointed out in the Introduction, for a two group comparison, both Goeman's and Hummel's methods are permutationally equivalent to Hotelling's T^2 under the assumption of equal variance and zero covariance across all genes. This holds true for our location test as well, using Euclidean distance in a balanced completely randomized design. Since permutational equivalence holds regardless of the distribution of the data, in the case of balanced completely randomized designs all three tests have equal power as permutation tests. In the moderately unbalanced case, with similar within group variability, simulations (not shown) show that our test is still nearly permutationally equivalent to the other two tests so that the power will be similar.

Our proposed location test is based on the matrix of pairwise distances between microarrays and is therefore related in a natural way to cluster analysis which applies an algorithm to the distance matrix to find clusters in the data. In experimental designs involving two predetermined groups of microarrays cluster analysis can be used as a graphical technique to see if the clustering algorithm "finds" the predetermined groups. Our proposed method can be thought of as a formal significance test of whether two predetermined groups form two distinct clusters. Thus our test is consistent with an intuitive visual display of the data. Of course applying our location test, or any significance test, to two groups that were discovered using a clustering algorithm would be circular and therefore invalid.

Author Contributions

Conceived and designed the experiments: DH DS. Performed the experiments: DH. Analyzed the data: DH. Contributed reagents/materials/analysis tools: DH PL. Wrote the paper: DH DS.

References

1. Chen JJ, Lee T, Delongchamp RR, Chen T, Tsai CA (2007) Significance analysis of groups of genes in expression profiling studies. Bioinformatics 23: 2104–2112.
2. Allison DB, Cui X, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. Nature Reviews Genetics 7: 55–65.
3. Goeman JJ, Bülmann P (2007) Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics 23: 908–987.
4. Khatri P, Drăghici S (2005) Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics 21: 3587–3595.
5. Nam D, Kim SY (2008) Gene-set approach for expression pattern analysis. Briefings in Bioinformatics 9: 189–197.
6. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC (2004) A global test for groups of genes: testing association with clinical outcome. Bioinformatics 20: 93–99.
7. Kong SW, Pu WT, Park PJ (2006) A multivariate approach for integrating genome-wide expression data and biological knowledge. Bioinformatics 22: 2373–2380.
8. Hummel M, Meister R, Mansmann U (2008) GlobalANCOVA: exploration and assessment of gene group effects. Bioinformatics 24: 78–85.
9. Hotelling H (1931) The generalization of student's ratio. Ann Math Statist 2: 360–378.
10. Pesarin F (2001) Multivariate Permutation Tests: With Applications to Biostatistics. Chichester, England: John Wiley & Sons LTD. pp 43.
11. Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap Chapman & Hall/CRC. pp 202–219.
12. Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27 Part 1: 209–220.
13. Berry KJ, Kvamme KL, Mielke Jr PW (1983) Improvements in the permutation test for the spatial analysis of the distribution of artifacts into classes. American Antiquity 48(3): 547–553.
14. Snedechor GW, Cochran WG (1989) Statistical Methods. Ames: Iowa State University Press. pp 177.
15. Fleiss JL (1986) The Design and Analysis of Clinical Experiments John Wiley & Sons. pp 2–4.
16. Rajicic N, Finkelstein DM, Schoenfeld DA, et al. (2006) Survival analysis of longitudinal microarrays. Bioinformatics 22: 2634–2649.
17. Samuel K, Johnson NL, editors (1982) Encyclopedia of Statistical Sciences, volume 1 John Wiley & Sons. pp 205–209.