A Discussion Concerning the Inclusion of Variety Effect when Analysis of Variance is Used to Detect Differentially Expressed Genes

Guri Feten, Are Halvor Aastveit, Lars Snipen and Trygve Almøy

Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O.Box 5003, N-1432 Ås, Norway.

Abstract: In microarray studies several statistical methods have been proposed with the purpose of identifying differentially expressed genes in two varieties. A commonly used method is an analysis of variance model where *only* the effect of interaction between variety and gene is tested. In this paper we argue that in addition to the interaction effects, the main effect of variety should simultaneously also be taken into account when posting the hypothesis.

Keywords: Hypothesis test, main and interaction effects, microarray experiment.

1. Introduction

The technology of DNA microarray gives us the opportunity to screen the expression levels for thousands of genes simultaneously. An introduction to this technology can be found in e.g. Nguyen et al. (2002). One of the important issues when measuring gene expression microarray is to search for genes which are Differentially Expressed (DE) in two different varieties. A gene is DE if the amount of mRNA is different in the two samples compared. The samples compared can be from different times of a biological process, different strains of a bacterium, or different conditions (e.g. treatment or disease).

Many statistical models and methods have been proposed in the search for DE genes. In this paper we will focus on the Analysis of Variance (ANOVA) model. This model for finding DE genes was introduced by Kerr et al.(2000, 2001, 2002), Wolfinger et al. (2001), and Kerr and Churchill (2001a, 2001b). A commonly used hypothesis for identifying DE genes is to test merely the effect of interaction between variety and gene. We argue that in addition to the interaction effects the main effect of variety should be included simultaneously in the hypothesis. The latter hypothesis is briefly discussed by Black and Doerge (2002) and Craig, Black and Doerge (2003), but not as the main topic of their papers. The hypothesis introduced by Kerr et al. is widely spread, and is still popular, even after the introduction of the hypothesis with the main effect included. Due to this, we do a more thorough discussion on the two hypotheses and discuss the consequences of choosing one of them in preference to the other one.

The outline of the paper is as follows. In Section 2 we state the model and two different hypotheses for detecting DE genes, one based on the interaction effect and the other based on both the interaction effect and the main effect simultaneously. In addition, we discuss the differences, both with new arguments and with arguments taken from Black and Doerge (2002) and Craig et al. (2003). Both hypotheses are tested on an example in Section 3, and in Section 4 we summarize the results obtained.

2. The Model

In previously published papers concerning the problem of detecting DE genes, different authors have applied ANOVA-models (Kerr et al. 2000, 2001, 2002, Wolfinger et al. 2001, Kerr and Churchill, 2001a, 2001b, Craig et al. 2003). The model suggested in Kerr et al. (2000) (from now on denoted Kerr) is

Correspondence: Guri Feten, Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway. Tel: +47 64 96 58 28; Fax: +47 64 96 59 01; Email: guri.feten@umb.no

$$y_{ijkg} = \mu + A_i + D_j + V_k + G_g \\ + (AG)_{ig} + (VG)_{kg} + \epsilon_{ijkg}, \qquad (1)$$

where y_{ijkg} is the expression level on the log scale for gene g of variety k measured on array i using dye j. Further, μ is the overall mean, A_i is the effect of the ith array, D_j is the effect of dye j, V_k is the effect of the kth variety, G_g is the effect of gene g, $(AG)_{ig}$ is the interaction effect between array i and gene g, $(VG)_{kg}$ is the interaction effect between variety k and gene g, and ϵ_{ijkg} is an error term independently distributed as $\epsilon_{ijkg} \sim N(0, \sigma^2)$. In this model D_j, V_k, G_g, and $(VG)_{kg}$ are assumed to be fixed, while A_i and $(AG)_{ig}$ are commonly assumed to be random.

There are sixteen possible effects, included interactions of all orders. With a dye-swap experiment each of the sixteen effects is completely confounded with one other effect, hence one effect is only possible to estimate assuming the confounding effect is zero. A table showing the pairs of confounded effects in a dye-swap experiment can be found in Kerr.

Additional effects, such as the interaction between dye and gene, $(DG)_{jg}$, could have been included in the model. However, these effects use degrees of freedom that should be used for estimating error. Omitting the effect of $(DG)_{jg}$ in the analysis of a dye-swap experiment will not alter the estimates of the other effects in the model, due to the orthogonality of the design.

The objective of the experiment is to identify genes that are differentially expressed in treatment and control samples. Previous papers state that genes with significant interaction between variety and gene are DE. Hence, the effect of interest in Model (1) is $(VG)_{kg}$ (Kerr et al. 2000, 2001, 2002, Wolfinger et al. 2001, Kerr and Churchill, 2001a, 2001b). According to these authors, a gene g is DE if and only if $(VG)_{1g} - (VG)_{2g}$ is non-zero. For fixed effects the restriction $\Sigma_{k=1}^2 (VG)_{kg} = 0$ is usually assumed, hence $(VG)_{1g} - (VG)_{2g} \neq 0$ corresponds to $(VG)_{kg} \neq 0$. This gives the following statement:

Statement 1
A gene g is DE if and only if $(VG)_{kg} \neq 0$.

The hypothesis of interest,

$$H_{0g}: (VG)_{kg} = 0, \qquad (2)$$

is tested with a t-test for each gene separately.

In this paper we will show that the study of the difference $(VG)_{1g} - (VG)_{2g}$ in some cases is not accurate enough, in fact it is not even correct, hence, Statement 1 is incorrect. If our interest is focused on whether a gene is DE in two varieties, we should rather examine the corresponding expected difference between the gene expressions in the two varieties. If we assume the restrictions $\Sigma_{k=1}^2 V_k = 0$ and $\Sigma_{k=1}^2 (VG)_{kg} = 0$, this is given by

$$E(\bar{y}_{\cdot\cdot 1g} - \bar{y}_{\cdot\cdot 2g}) = (V_1 - V_2) + ((VG)_{1g} - (VG)_{2g}) \\ = 2V_1 + 2(VG)_{1g},$$

where $\bar{y}_{\cdot\cdot kg} = \frac{1}{2a} \Sigma_{i=1}^a \Sigma_{j=1}^2 y_{ijkg}$ for $k = 1, 2$. A gene is DE if there is an effect of variety, either main effect, interaction effect, or both. This gives the following statement:

Statement 2
A gene g is DE if and only if $(V_1 + (VG)_{1g}) \neq 0$.

The hypothesis of interest,

$$H_{0g}: V_1 + (VG)_{1g} = 0, \qquad (3)$$

is carried out with a t-test for each gene separately. The two hypotheses are equal if and only if $V_1 = 0$. The hypothesis in (3) is presented in Black and Doerge (2002) as an alternative hypothesis to the hypothesis in (2).

The differences between Statement 1 and 2 are illustrated in Figure 1. According to both statements both genes in situation 1 are DE. In situation 2, Statement 1 claims both genes to be DE, while Statement 2 claims gene one to be Equally Expressed (EE) and gene two to be DE. In situation 3 both genes are EE according to Statement 1, while both genes are DE according to Statement 2. According to both statements both genes are EE in situation 4.

If only the interaction terms are used to test for DE genes, the conclusions are dependent on the choice of genes printed on the array. Adding a new gene to the array, or removing a gene from the array, may change the level of interaction, and hence the conclusions about which genes are DE.

A necessary assumption if Statement 1 should be preferred, is that the majority of the genes are
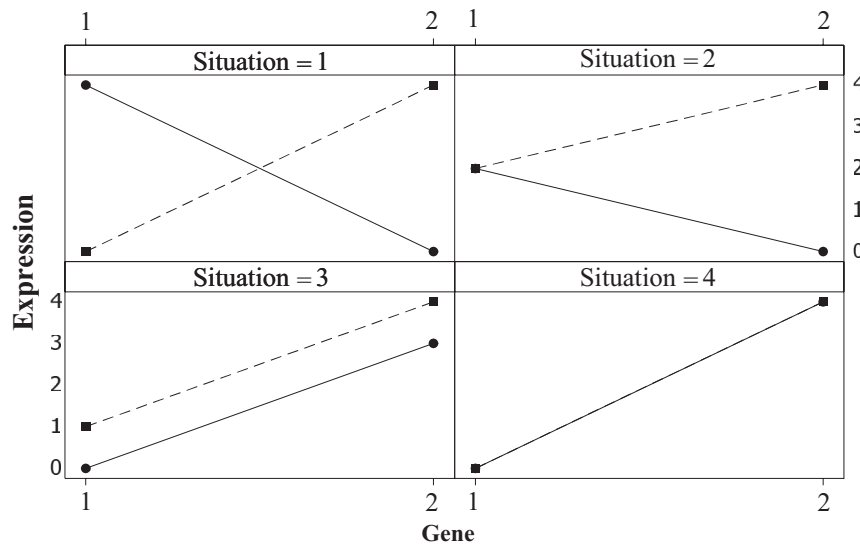
Figure 1. Different situations for main effect and interaction between variety and gene. Variety 1 (——●——) and Variety 2 (– – ■ – –).

equally expressed, and that differentially expressed genes, either up- or down-regulated, will balance each other out. Then there is no true effect of variety, and the corresponding term can be excluded from the hypotheses. However, differentially expressed genes occur more frequently than previously assumed (van de Peppel et al. 2003).

Both Black and Doerge (2002) and Craig et al. (2003) suggest that Statement 2 is to prefer when there is a true difference between the varieties averaged over all genes, while Statement 1 is to prefer when the variety effect is an artifact of the experimental process, however, this is usually difficult to decide. For dye-swap experiments the effect of variety and the interaction between array and dye are confounded, while for reference designs the effect of variety confounds with the effect of array, hence the effect of variety cannot be separated from experimental effects.

To get an idea about any possible effect of interaction between array and dye (or with reference designs, the effect of dye), a pilot study with self-self hybridization can be performed. Samples from one of the varieties are divided in two, and then labeled with Cy5 and Cy3, respectively. Hence there is no true variety effect, and any estimated effect is interaction of array and dye. If this interaction effect seems large, the hypothesis in (2) should be applied, while if the interaction effect seems small or ignorable, the hypothesis in (3) should be applied.

As also stated in Kerr (2003) a replacement of the term V_k in Model (1) with the confounding interaction effect between array i and dye j, $(AD)_{ij}$, gives the following model

$$y_{ijkg} = \mu + A_i + D_j + (AD)_{ij} + G_g \\ + (AG)_{ig} + (VG)_{kg} + \epsilon_{ijkg},$$

which clarifies that the term represents an effect of the experiment and not an effect of variety.

3. Example

In order to study the consequences of Statement 1 and 2, we have reanalyzed the data used in Kerr. Samples of mRNA from human liver tissue and from muscle tissue were compared. The analyzed data contains 1286 clone identifiers representing 1274 genes. The purpose of the experiment was to test whether the clone identifiers (from now on called genes) are differentially expressed between the two types of tissues. The experiment was designed as a latin square (i.e. dye-swap). Model (1) was fitted to the data, and the analysis of variance is given in Table 1, which is an extended version of Table 3 in Kerr. The analysis is performed assuming array effect fixed as done in Kerr. The variety effect and the interaction effect of variety × gene seem to be significant, but due to lack of normality the p-values in the table are only approximate values.

In order to see the difference between the hypotheses in (2) and (3), we carried out the tests in both situations. Due to the deviation from

Table 1. Analysis of variance for the microarray experiment.

Source	DF	SS	MS	F	P
Array	1	92.34	92.34	1433.87	<0.001
Dye	1	0.74	0.74	11.56	0.001
Variety	1	2.97	2.97	46.11	<0.001
Gene	1285	1885.89	1.47	22.79	<0.001
Array × Gene	1285	160.01	0.12	1.93	<0.001
Variety × Gene	1285	1357.28	1.06	16.40	<0.001
Error	1285	82.75	0.0644		
Total	5143	3581.99			

normally distributed errors, we (and Kerr) applied a bootstrap analysis of the residuals (Efron and Tibshirani, 1986), instead of the usual confidence intervals which are based on the assumption of normally distributed errors. We produced a new set of observations by resampling, with replacement, from the original residuals. The new observations were then given by

$$y^*_{ijkg} = \hat{\mu} + \hat{A}_i + \hat{D}_j + \hat{V}_k + \hat{G}_g$$
$$+ (\widehat{AG})_{ig} + (\widehat{VG})_{kg} + \epsilon^*_{ijkg},$$

where the parameters are the estimated parameters from the original fit of the model, and ϵ^*_{ijkg} are drawn independently from $\sqrt{4m/(m-4)}\,\hat{\epsilon}$, where m is the number of genes and $\hat{\epsilon}$ are the original residuals. We rescale the original residuals to produce a distribution with the same variance as the true residuals (Kerr). Based on 1000 bootstrap data sets we computed 95% confidence intervals for the differences

$$(VG)_{1g} - (VG)_{2g},$$

and for the differences

$$E(\bar{y}_{\cdot\cdot 1g} - \bar{y}_{\cdot\cdot 2g}) = (V_1 - V_2) + ((VG)_{1g} - (VG)_{2g}).$$

If the confidence interval does not contain zero, the gene is DE, and either the liver tissue or the muscle tissue have greater expression, depending on the magnitude of the difference. We have not considered the problem of multiple comparisons, and hence the problem of adjusted p-values, neither have Kerr. The results are shown in Table 2. The tests of the two hypotheses conclude differently for 3.1% of the genes. If testing the hypotheses in (2), 17 genes would wrongly (compared to tests of the hypotheses in (3)) be denoted as DE, while 23 genes would wrongly be denoted as EE.

4. Conclusion

Methods to detect DE genes have been introduced and applied to microarray data. Among others, Kerr states that "changes in gene expression across experimental samples are captured in the variety × gene interaction terms of the ANOVA model". If this is correct, the interaction terms can be studied to test the hypothesis of differential expression for individual genes. In this paper we have studied the use of analysis of variance models and hypotheses based on these models, and possible weaknesses of these hypotheses.

Table 2. Number of genes categorized to DE (either Liver < Muscle or Liver > Muscle) or EE by tests of the hypotheses in (2) and (3) respectively.

$(VG)_{1g} - (VG)_{2g}$	$(V_1 - V_2) + ((VG)_{1g} - (VG)_{2g})$			
	Liver < Muscle	Liver = Muscle	Liver > Muscle	
Liver < Muscle	374	0	0	29.1%
Liver = Muscle	23	659	0	53.0%
Liver > Muscle	0	17	213	17.9%
	30.9%	52.6%	16.5%	1286

We have shown that in situations where there is effect of variety, it is not enough to examine the interaction between variety and gene. This interaction effect will only measure the deviation from the average effect of variety (Black and Doerge, 2002). If the effect of variety increases, more genes become incorrectly denoted as either DE or EE. In situations with no proved significant effect of variety, it is enough to study only the interactions. However, it is not to be recommended, since the test of variety effect may be incorrectly accepted.

References

Black, M.A. and Doerge, R.W. 2002. Calculation of the minimum number of replicate spots required for detection of significant gene expression fold change in microarray experiments. *Bioinformatics,* 18:1609–1616.

Craig, B.A., Black, M.A. and Doerge, R.W. 2003. Gene expression data: The technology and statistical analysis. *Journal of Agricultural, Biological, and Environmental Statistics,* 8:1–28.

Efron, B. and Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science,* 1:54–77.

Kerr, M.K. 2003. Linear models for microarray data analysis: Hidden similarities and differences. *Journal of Computational Biology,* 10:891–901.

Kerr, M.K. and Churchill, G.A. 2001a. Statistical design and the analysis of gene expression microarray data. *Genetical Research,* 77:123–128.

Kerr, M.K. and Churchill, G.A. 2001b. Experimental design for gene expression microarrays. *Biostatistics,* 2:183–201.

Kerr, M.K., Leiter, E.H., Picard, L. and Churchill, G.A. 2001. Analysis of a designed microarray experiment. Proceedings of the IEEE-Eurasip Nonlinear Signal and Image Processing Workshop, June 3–6, 2001.

Kerr, M.K., Martin, M. and Churchill, G.A. 2000. Analysis of variance for gene expression microarray data. *Journal of Computational Biology,* 7:819–837.

Kerr, M.K., Afshari, C.A., Bennett, L., Bushel, P., Martinez, J., Walker, N.J. and Churchill, G.A., 2002. Statistical analysis of a gene expression microarray experiment with replication. *Statistica Sinica,* 12:203–217.

Nguyen, D.V., Arpat, A.B., Wang, N.Y. and Carroll, R.J. (2002). DNA microarray experiments: Biological and technological Aspects. *Biometrics,* 58:701–717.

van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D. and Holstege, F.C.P. 2003. Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO reports,* 4:387–393.

Wolfinger, R.D., Gibson, G., Wolfinger, E.D., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C. and Paules, R.S. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology,* 8:625–637.