

Gene expression

Increased power of microarray analysis by use of an algorithm based on a multivariate procedure

K. Krohn¹, M. Eszlinger², R. Paschke², I. Roeder³ and E. Schuster^{3,*}¹Interdisciplinary Center for Clinical Research Leipzig, University of Leipzig, Inselstrasse 22, 04103 Leipzig, Germany, ²III. Medical Department, University of Leipzig, Ph.-Rosenthal-Strasse 27, 04103 Leipzig, Germany and³Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Haertelstrasse 16-18, 04107 Leipzig, Germany

Received on February 11, 2005; revised and accepted on July 1, 2005

Advance Access publication July 5, 2005

ABSTRACT

Motivation: The power of microarray analyses to detect differential gene expression strongly depends on the statistical and bioinformatical approaches used for data analysis. Moreover, the simultaneous testing of tens of thousands of genes for differential expression raises the ‘multiple testing problem’, increasing the probability of obtaining false positive test results. To achieve more reliable results, it is, therefore, necessary to apply adjustment procedures to restrict the family-wise type I error rate (FWE) or the false discovery rate. However, for the biologist the statistical power of such procedures often remains abstract, unless validated by an alternative experimental approach.

Results: In the present study, we discuss a multiplicity adjustment procedure applied to classical univariate as well as to recently proposed multivariate gene-expression scores. All procedures strictly control the FWE. We demonstrate that the use of multivariate scores leads to a more efficient identification of differentially expressed genes than the widely used MAS5 approach provided by the Affymetrix software tools (Affymetrix Microarray Suite 5 or GeneChip Operating Software). The practical importance of this finding is successfully validated using real time quantitative PCR and data from spike-in experiments.

Availability: The R-code of the statistical routines can be obtained from the corresponding author.

Contact: Schuster@imise.uni-leipzig.de

INTRODUCTION

High-density oligonucleotide microarray technology is increasingly used for gene-expression profiling (Gershon, 2002), to define a group of genes with differential expression between a variety of experimental conditions. The power of such analyses depends not only on the quality of array design and production, but also on the statistical and bioinformatical approaches used to analyse the data. Indeed, the application of different mathematical algorithms can influence enormously the outcome of microarray data analysis, e.g. by having different statistical power to detect significant differentially expressed genes. Many practicing biologists are unaware of the extent of this problem. Therefore, it is important not just to develop strategies with optimal theoretical properties, but to demonstrate their practical relevance by additional experimental tests.

Affymetrix GeneChips have become particularly prevalent within the field of microarray gene-expression analysis and the results of

GeneChip experiments attract the attention of a wide spectrum of life science researchers (Harrington *et al.*, 2000). The common practice of analysing differential gene expression from GeneChip data is based on normalized hybridization signals generated with the Affymetrix Microarray Suite (MAS) 5 software algorithms. This method returns a single expression value per gene, condensing the information from hybridizing up to 18 pairs of perfect match (PM) and mismatch (MM) oligonucleotides (known collectively as a probe set) to complementary mRNA. The subsequent simultaneous testing of tens of thousands of genes for differential expression raises the ‘multiple testing problem’, increasing the probability of obtaining false positive test results (Westfall and Young, 1993). To achieve more reliable results, it is, therefore, necessary to restrict the family-wise type I error rate (FWE) or the false discovery rate (FDR) by application of adjustment procedures (Reiner *et al.*, 2003; Storey and Tibshirani, 2003). However, classical approaches, such as the Bonferroni correction of genewise *t*-test results are highly conservative and their application leads to a dramatic loss of statistical power (i.e. a high number of false negative results), when testing thousands of genes for differential expression.

In contrast to MAS5 and in agreement with alternative approaches to microarray data analysis (Li and Wong, 2001; Irizarry *et al.*, 2003), we propose a procedure that employs the complete multivariate information from all PM oligonucleotides complementary to an individual transcript. Our strategy, which is based on the theory of spherical distributions (Läuter *et al.*, 1998), strictly maintains FWE at a prespecified level α (Westfall *et al.*, 2004) and enables more efficient detection of differentially expressed genes than do approaches based on conventional expression scores (e.g. MAS5). To demonstrate the practical importance of this finding, we successfully validated differential gene expression detected with the multivariate procedure using real time quantitative PCR and data from spike-in experiments.

METHODS**Multiplicity adjustment procedure
[Westfall–Kropf–Finos (WKF) procedure]**

The proposed methodology is applicable to situations of two dependent or independent samples as well as to one-sample situations (Läuter *et al.*, 1996). Please note that the case of two dependent samples can be treated as a one-sample problem considering measurement differences within

*To whom correspondence should be addressed.

individuals. We will, therefore, demonstrate the methodology for a dependent and for an independent sample data set (for details see below). However, in this section we restrict ourselves to the case of two dependent samples for reasons of brevity.

Let us consider n independent, identically normal distributed k -dimensional sample vectors \mathbf{z}_j (with components z_{ji} , $i = 1, \dots, k$;

$$j = 1, \dots, n) \text{ with expectation } \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}. \text{ Here, } n \text{ represents the number}$$

of individual specimens for which differences in the expression of k probe sets (representing certain genes) between two types of tissue probes have been analysed. To test the local hypotheses $H_i : \mu_i = 0$ for $i = 1, \dots, k$ (no difference in gene expression for all genes), Westfall *et al.* (2004) introduced the following procedure:

- (1) Calculate the P -values P_i ($i = 1, \dots, k$) using unadjusted one-sample t -tests for the z_{ji} . ($j = 1, \dots, n; i = 1, \dots, k$)
- (2) For each variable (e.g. probe set), determine the sum of square values $w_i = \sum_{j=1}^n z_{ji}^2$ ($i = 1, \dots, k$) and the weights $g_i = w_i^\eta$ for a fixed value $\eta \geq 0$. Calculate the weighted P -values $Q_i = P_i/g_i$ and sort the variables according to $Q_{i_1} \leq Q_{i_2} \leq \dots \leq Q_{i_k}$, which gives the following order: $Q_{(1)} \leq Q_{(2)} \leq \dots \leq Q_{(k)}$. Define the index sets $S_j = \{i_j, i_{j+1}, \dots, i_k\}$ ($j = 1, \dots, k$). The ordered hypotheses $H_{(j)}$ ($j = 1, 2, \dots$) are rejected as long as $Q_{(j)} \leq \frac{\alpha}{\sum_{m \in S_j} g_m}$.

Stop at the first j yielding a value $Q_{(j)}$ which does not meet this inequality. Westfall, Kropf and Finos (Westfall *et al.*, 2004) showed that this procedure maintains the FWE α if the \mathbf{z}_j follows a multivariate normal distribution. The procedure is equivalent to the classical Bonferroni–Holm procedure for the specific parameter choice $\eta = 0$. Also the other special case, $\eta \rightarrow \infty$, has already been discussed in detail (Kropf, 2000; Kropf and Läuter, 2002). In the following paragraphs, this procedure is considered for arbitrary $\eta \geq 0$ and is applied to multivariate PM-based scores which are left-spherically (Läuter *et al.*, 1998) rather than normally distributed. It can be shown that this procedure (referred to as WKF procedure) maintains the FWE also in this more general setting (Schuster *et al.*, 2004).

Gene-expression scores

MAS 5 The algorithm calculates a weighted mean for the probe set using signal intensities of PM and MM oligonucleotides in a one-step Tukey’s biweight estimate. Signal intensity of a probe pair (PM and MM) is estimated by taking the log of the PM intensity minus stray signal calculated from mismatch intensities (Affymetrix technical note: ‘Statistical Algorithms Reference Guide’, www.affymetrix.com).

MDP This procedure uses a robust version of a two-way analysis of variance (considering chip and probe effects) to estimate the expression value for each individual gene (Izarrary *et al.*, 2003).

Multivariate scores Let us assume that the PM oligonucleotides follow a multivariate normal distribution with expectations possibly different from zero. k is the number of probe sets and p_i the PM number of probe set i . The total PM number is, therefore, given by $p = \sum_{i=1}^k p_i$. The column vector containing all p_i PMs of probe set i and individual j is denoted by \mathbf{x}_{ji} . With \mathbf{x}_j we now denote the n independent, identically normal-distributed p -dimensional sample vectors

$$\mathbf{x}_j = \begin{pmatrix} \mathbf{x}_{j1} \\ \vdots \\ \mathbf{x}_{jk} \end{pmatrix} \sim N_p(\mu, \Sigma) \quad (j = 1, \dots, n),$$

with expectation $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}$ and covariance matrix $\Sigma = \begin{pmatrix} \Sigma_{11} & \dots & \Sigma_{1k} \\ \vdots & \ddots & \vdots \\ \Sigma_{k1} & \dots & \Sigma_{kk} \end{pmatrix}$.

\mathbf{x}_j is, therefore, a column vector representation of the individual row j of the $n \times p$ dimensional data matrix \mathbf{X} . The total product sum matrix is given by

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \dots & \mathbf{W}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{k1} & \dots & \mathbf{W}_{kk} \end{pmatrix} = \mathbf{X}'\mathbf{X}. \text{ And the local hypotheses are } H_i : \mu_i =$$

0 ($i = 1, \dots, k$) at a FWE level α_{FWE} .

To create multivariate PM-based scores of the expression intensities we form weighting vectors, which depend only on the total product sum matrix \mathbf{W} of the data. Within this dependency postulation, different choices of weighting vectors are possible. Four options are described briefly below [for details see (Läuter *et al.*, 1996)].

The weighting vectors \mathbf{c}_i ($i = 1, \dots, k$) are obtained for the

- (1) non-standardized principal component (NPC) test as the eigenvectors \mathbf{c}_i ($i = 1, \dots, k$) of the largest eigenvalue of the eigenvalue problems $\mathbf{W}_{ii}\mathbf{c}_i = \lambda_i\mathbf{c}_i$ with $\mathbf{c}_i'\mathbf{c}_i = 1$;
- (2) standardized principal component (SPC) test as the eigenvectors \mathbf{c}_i ($i = 1, \dots, k$) of the largest eigenvalue of the eigenvalue problems $\mathbf{W}_{ii}\mathbf{c}_i = \text{Diag}(\mathbf{W}_{ii})\mathbf{c}_i$ with $\mathbf{c}_i'\text{Diag}(\mathbf{W}_{ii})\mathbf{c}_i = 1$ ($i = 1, \dots, k$);
- (3) covariance sum (CS) test as $\mathbf{c}_i = [\text{Diag}(\mathbf{W}_{ii})]^{-1}\mathbf{W}_{ii}[\text{Diag}(\mathbf{W}_{ii})]^{-1/2}\mathbf{1}_{p_i}$;
- (4) standardized sum (SS) test as $\mathbf{c}_i = [\text{Diag}(\mathbf{W}_{ii})]^{-1/2}\mathbf{1}_{p_i}$.

Please note that all \mathbf{c}_i are dependent on \mathbf{W}_{ii} only.

To ensure that probe sets with balanced increased and decreased oligonucleotide measurements \mathbf{x}_{ji} are not counted as differentially expressed, the following absolute and standardized weighting vectors (only depending on \mathbf{W}_{ii}) are used:

$$\mathbf{d}_{ji} = \frac{|c_{il}|}{\sum_{l=1}^{p_i} |c_{il}|}, \quad l = 1, \dots, p_i, \quad i = 1, \dots, k.$$

Now, the p_i -dimensional data vectors of PMs for probe set i and individual j are subsumed into the score: $z_{ji} = \mathbf{d}_{ji}'\mathbf{x}_{ji}$, which is unequal to zero with probability one.

It can be demonstrated that the scores z_{ji} are left-spherically distributed whenever the original data are multivariate normal.

It can be proven (Schuster *et al.*, 2004) that the WKF procedure for the left-spherically distributed scores z_{ji} keeps a given FWE under the assumption of multivariate normal distributed \mathbf{x}_{ji} . It should be noted that this also holds for the two-sample situation with weights determined by an appropriate modification of the matrix \mathbf{W} (Schuster *et al.*, 2004).

SAMPLE DATASETS

Two sample datasets have been included in this study to test the power of the different algorithms.

The first set (dependent samples) contains patient data from a gene-expression project which characterizes molecular events in thyroid tumour tissue (Eszlinger *et al.*, 2004). Here, the description of differentially expressed genes will further our understanding of the impaired thyroid epithelial cell signalling that eventually leads to thyroid neoplasia and is, therefore, relevant to both diagnosis and therapy of thyroid tumours. Gene-expression analysis using Affymetrix GeneChips was performed on a set of 15 autonomously functioning thyroid nodules (AFTNs) and matching normal surrounding tissue taken during thyroid surgery. Experimental procedures for the set of 30 Affymetrix GeneChip experiments follow the manufacturer’s instructions (Affymetrix, Santa Clara, CA, USA) and have been recently described (Eszlinger *et al.*, 2004). In addition to microarray experiments, aliquots of total RNA preparations were also used to quantify selected genes in a real-time reverse transcribed (RT)–PCR reaction using a LightCycler (Roche, Mannheim, Germany) as previously described (Eszlinger *et al.*, 2001, 2004).

Table 1. Reevaluation of transcripts (defined by the Affymetrix probe set ID and the gene symbol) obtained as differentially expressed by the test procedures MAS5, MDP or the variants of the WKF procedure, NPC and SPC [$\eta = 1$, as suggested by Schuster *et al.* (2004) for cases without preliminary information]

Probe set ID	Gene symbol	GeneChip analysis						Real time PCR				
		SLR (MAS5)		WKF statistics				LR		Statistics		
		Mean	SEM	MAS5 $\eta = 1$	MDP $\eta = 1$	NPC $\eta = 1$	SPC $\eta = 1$	mean	SEM	P-value unadjusted	Bonferoni-Holm	
31812_at	GMPR	0.0	0.1	N	N	S	N	0.0	0.4	0.9889	N	
384_at	PSMB10	0.1	0.1	N	N	S	N	-0.8	0.6	0.1961	N	
37653_at	COL18A1	0.1	0.2	N	N	S	S	0.7	0.2	0.0061	S	
38451_at	UQCR	0.6	0.1	N	S	S	S	0.6	0.2	0.0126	S	
41206_r_at	COX6A1	0.6	0.1	N	S	S	S	0.7	0.3	0.0184	S	
37033_s_at	GPX1	0.7	0.1	N	S	S	S	0.7	0.3	0.0215	S	
37952_at	ITGB3	0.7	0.1	N	S	S	S	0.9	0.3	0.0042	S	
37906_at	LTBP2	-1.1	0.2	N	S	S	S	-1.0	0.3	0.0017	S	
343_s_at	ENPP1	-1.1	0.2	N	S	S	S	-1.9	0.3	0.0001	S	
38059_g_at	DPT	-1.9	0.4	N	S	S	S	-2.4	0.5	0.0003	S	
1736_at	IGFBP6	-2.1	0.4	N	S	S	S	-1.6	0.2	0.00001	S	
35649_at	CDO1	-2.2	0.4	N	S	S	S	-2.3	0.3	0.000001	S	
39757_at	SDC2	-1.1	0.1	S	S	S	S	-1.2	0.2	0.0003	S	
39066_at	MFAP4	-2.4	0.3	S	N	S	N	-2.4	0.5	0.0008	S	
36455_at	COL9A3	3.4	0.5	S	S	S	S	3.3	0.7	0.0003	S	
38057_at	DPT	-1.9	0.3	S	S	S	S	-2.4	0.5	0.0003	S	
33249_at	NR3C2	-0.9	0.2	S	N	S	N	-0.8	0.2	0.0032	S	
36167_at	ATP6F	0.5	0.0	S	S	S	S	0.4	0.2	0.0144	S	
31966_at	DIO1	2.0	0.2	S	S	S	S	2.0	0.3	0.00001	S	
36681_at	APOD	-3.5	0.3	S	S	S	S	-3.1	0.4	0.000002	S	
35928_at	TPO	1.7	0.2	S	S	S	S	1.4	0.3	0.0007	S	
926_at	MT1G	1.3	0.2	S	S	S	S	1.6	0.3	0.0001	S	
578_at	RAG2	1.6	0.3	S	N	N	N	2.7	0.3	0.000001	S	
672_at	SERPINE1	1.5	0.2	S	N	N	N	1.2	0.3	0.0005	S	

Shown are data for GeneChip and real-time RT-PCR experiments for a selection (for details see 'sample dataset') of 24 differentially expressed probe sets ($\alpha_{FWE} = 0.10$) out of the 12 625 probe sets in 30 specimens (tumours and matched normal tissues from 15 patients). Results of the GeneChip experiments are given as mean and SEM of the signal log ratio (SLR) for the expression in tumour compared with normal tissue calculated with the Affymetrix MAS5 software. Similarly, for real time PCR the \log_2 of the ratio (LR) for the gene expression in matched tissue pairs is given as mean and SEM. Statistical analysis was done according to Bonferoni-Holm ($\alpha_{FWE} = 0.10$; $k = 24$). N, not significant; S, significant.

Briefly, 2 μg of total RNA was reverse transcribed using SuperScript™ II Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) primed with oligo-dT according to the manufacturer's instructions. Optimal PCR reactions for all investigated genes were established using the LightCycler—DNA Master SYBR Green I Kit (Roche, Mannheim, Germany) according to the manufacturer's instructions: annealing temperatures and MgCl_2 concentrations were optimized to create a one-peak-melting curve. Exact conditions and the nucleotide sequences of the PCR primers are available on request. PCR fragments were cloned into the pGEM-T vector (Promega, Madison, WI, USA). Dilutions of these plasmid preparations containing known copy numbers were used as calibration curves for each template. In addition, quantification of tumour and matched normal tissues from 15 patients were performed in the same run. To normalize for differences in the amount of cDNA added to the reactions, quantification of GAPDH and β -actin was performed as an endogenous control. The differential expression of the investigated genes was calculated as the ratio tumour/surrounding tissue and expressed as \log_2 .

From three pools, 24 probe sets for quantitative RT-PCR (Table 1) were selected.

- (1) Ten probe sets detected as differentially expressed with both MAS5 and NPC procedures were randomly selected.

- (2) Six probe sets detected only as differentially expressed with the NPC algorithm were randomly selected. Additionally, we picked the three samples with the highest/lowest score.
- (3) From the sets detected only with MAS5, we picked the two probe sets that were in this pool.

Second, we have compared the different approaches using a latin square dataset (independent samples) of spike-in experiments (Human Genome U133 dataset) that is available from Affymetrix (https://www.affymetrix.com/support/technical/sample_data/datasets.affx). The dataset consists of 42 GeneChip arrays based on three technical replicates of 14 experimental settings (E1–E14). In each setting 39 of 42 transcripts are spiked into a complex human RNA background at concentrations ranging from 0.125 to 512 pM. A different combination of 3 of the 42 transcripts is left out in each setting. Except for transcripts that are left out, the concentration of the spike doubles from one to the next setting. Therefore, when comparing, for example, the three replicates of E1 to the three replicates of E2 we are able to validate the doubling of 36 transcripts (42 transcripts minus the three left out in E1 and minus the three left out in E2). As a result of this design, in each comparison of two experiments (E1 and E2, E2 and E3, ..., E14 and E1; altogether 14 comparisons)

a true 2-fold difference was present for 36 transcripts, whereas 22 258 transcripts were unchanged.

PROGRAMMING

The statistical results were obtained using the statistical programming environment R (Ihaka and Gentleman, 1996) including the BioConductor Package (www.bioconductor.org). Our analysis is based on AffyBatch objects obtained from the raw data (Affymetrix cel-file level) by application of quantile normalization without background correction. For the sake of comparison of the proposed multivariate approach with 'classical' methods we used summarized expression scores per probe set calculated by the R-method `expresso` with options `mas5` (MAS5) and the combination median-polish/pmonly (MDP). To approximately accommodate the criteria of multivariate normality and variance homogeneity, all expression values (individual PMs as well as expression scores) were expressed as \log_2 . The expression difference of a specific gene within individual patients (trait versus surrounding tissue) is expected to be zero, if its expression is not influenced by the trait. Therefore, the difference of the logarithms is tested against zero. The significance level has been set to $\alpha_{\text{FWE}} = 0.10$ in all procedures because of using FWE control which is a rather conservative criterion. It should be emphasized, that to strictly ensure the FWE criteria within the WKF procedure, the choice of η has to be made beforehand. If there is no prior knowledge from comparable studies, we recommend to use $\eta = 1$ (see also Schuster *et al.*, 2004).

The proposed statistical procedures as well as the thyroid tumour sample data are available within the data warehouse of the Interdisciplinary Centre for Bioinformatics Leipzig (<http://www.izbi.de/GEWARE>; see 'Public user groups'; procedures performed with 'multivariate expression analysis' under 'expression analysis').

RESULTS AND CONCLUSION

In the analysis of a set of 30 Affymetrix GeneChip experiments (Eszlinger *et al.*, 2004) the WKF multiplicity adjustment procedure (Methods section) returns the number of differentially expressed genes determined as significant for a choice of different values of a procedure specific parameter η (Fig. 1). To illustrate the advantages of the proposed multivariate approach, the WKF multiplicity adjustment procedure was applied to two cases which employ average probe set expression scores [MAS5 and the BioConductor routine medianpolish/pmonly (MDP) in Methods Section] and to four versions of true multivariate scores (NPC, SPC, CS and SS) as described above. The results for the thyroid tumour dataset show that the proposed total PM-based multivariate procedures as well as the MDP procedure are consistently superior to the conventional MAS5 approach in the sense of finding more significant probe sets without violating the FWE criteria for $\eta \leq 2$.

Regarding the proposed total PM-based multivariate procedures, this superiority is attributable to the incorporation of the total multivariate information contained in the individual PM oligonucleotides, i.e. our method is using individually (genewise) estimated weights to calculate expression values by weighted averaging procedures of the individual oligonucleotide measurements. In contrast, MAS5 is summarizing the oligonucleotide information of each probe set into expression scores by means of one fixed averaging procedure.

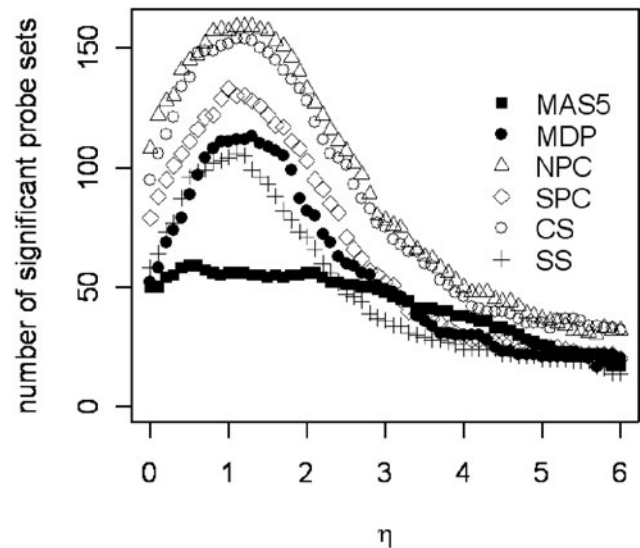


Fig. 1. Results of the WKF procedure. The numbers of significantly ($\alpha_{\text{FWE}} = 0.10$; $n = 12\,625$ probe sets) differentially expressed probe sets obtained by the application of the WKF procedure depending on parameter η . The different symbols correspond to the methods used. Herein, NPC test, CS test, SPC test and SS test are different versions of multivariate test scores applied within the new total PM-based procedure. Control results were obtained by the application of the multiple WKF procedure to expression score based t -tests constructed by the medianpolish/pmonly (MDP) and `mas5` (MAS5) options within the BioConductor-Routine `expresso`.

Furthermore, Figure 1 shows the WKF procedure to be better than the widely used Bonferroni–Holm procedure (which coincides with the WKF procedure using $\eta = 0$) for all choices of $0 < \eta < 2$. Comparing the different versions of multivariate scores, shows the CS test and the NPC test to be of similar quality, whereas the SPC test is noticeably inferior. Among the two expression score-based procedures the BioConductor routine MDP is superior to MAS5 at $\eta < 3$. This indicates that the procedure which MAS5 uses to summarize expression scores for the individual oligonucleotides of a probe set (Tukey's biweight) apparently leads to a loss of statistically useful information.

To assess the consistency of the results between the different approaches (individual oligo-based versus summarized expression score-based), we consider here the NPC and the MAS5 procedures at $\eta = 1$ (Table 1). The MAS5 procedure identified 56 probe sets with significant differential expression, whereas the NPC procedure identified 157, including 54 of the 56 detected by MAS5 and 103 detected uniquely by NPC. Only two probe sets were identified by MAS5 and not by NPC.

To validate the statistical results by an additional experimental approach, we tested subgroups of the respective genes or transcripts with real time quantitative RT-PCR (Eszlinger *et al.*, 2004). Strikingly, within the group of differentially expressed genes detected only by the NPC test, we were able to experimentally validate differential expression of 10 of 12 transcripts (Table 1). Moreover, all 10 probe sets from the group of genes common to the NPC and MAS5 procedure and the two probe sets that escape detection with the NPC show differential expression in real time quantitative PCR experiments.

Table 2. Results of the WKF procedure applied to data from spike-in experiments (for details see 'sample dataset')

Procedure	Percentage detection of true two fold different transcripts	Percentage detection of true negatives
CS	45.2	0.13
NPC	40.7	0.18
MDP	34.7	0.04
MAS	6.7	0.02
SS	50.0	0.06
SPC	52.6	0.06

The results of the first column have been generated as follows: First, we calculated the number of significantly ($\alpha_{FWE} = 0.10$; three versus three replicates) differentially expressed transcripts, with respect to the 36 cases with a true 2-fold difference (spiked-in transcripts that have a true 2-fold difference, 36 probe sets) for all 14 individual experiments. Second, we derive the average of these numbers. Third, the average is divided by 36 (number of transcripts with true 2-fold difference) and multiplied by 100. This leads to the percentage, which is given in the first column of the table. Calculation of the second column is similar. Only the 36 spiked-in transcripts are replaced by the 22 258 transcripts not different between the experiments.

We, therefore, conclude that the multivariate procedure is a more powerful means of detecting differentially expressed genes from microarray data than the standard MAS5 analysis.

This conclusion is also drawn when analysing data from spike-in experiments (Table 2). The percentage of detecting a transcript with a true 2-fold difference for all versions of multivariate procedures ranges from 40.7 to 52.6%. This is at least six times higher compared with MAS5 (6.7%). Moreover, in this dataset all different versions of the multivariate procedure show a better performance than the MDP algorithm, which detects a true 2-fold difference in $\sim 34.7\%$ of cases. We have to admit that the CS and the NPC procedure show a slight increase in false positive detection compared with MDP. This, however, might be of less practical importance, because the false positive rates of all procedures are very small (0.02–0.18 %). These false positives might also reflect cross-hybridization effects of the spiked transcripts which can partially explain their observed levels. When comparing the performance of the different versions of the multivariate procedure in the analysis of the two datasets, it becomes clear that it is currently not possible to specify an order of superiority between them.

Concluding from these analyses, we are recommending to avoid the use of the MAS5 procedure for identification of differentially expressed genes and present more powerful alternative approaches.

ACKNOWLEDGEMENTS

The authors thank M. Cross (IZKF Leipzig) and J. Läuter (IZBI Leipzig) for a critical reading of the manuscript and discussion of the project. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG/Pa423/10-1), the Interdisciplinary Center for Clinical Research (IZKF) at the University of Leipzig (project Z16-CHIP2) and the Deutsche Krebshilfe (project 106542). Microarray analysis was done at the IZKF Leipzig core facility. K.K. is supported by IZKF Leipzig, (project Z03).

Conflict of Interest: none declared.

REFERENCES

- Eszlinger, M. *et al.* (2004) Gene expression analysis reveals evidence for inactivation of the TGF-beta signaling cascade in autonomously functioning thyroid nodules. *Oncogene*, **23**, 795–804.
- Eszlinger, M. *et al.* (2001) Complementary DNA expression array analysis suggests a lower expression of signal transduction proteins and receptors in cold and hot thyroid nodules. *J. Clin. Endocrinol. Metab.*, **86**, 4834–4842.
- Gershon, D. (2002) Microarray technology: an array of opportunities. *Nature*, **416**, 885–891.
- Harrington, C.A. *et al.* (2000) Monitoring gene expression using DNA microarrays. *Curr. Opin. Microbiol.*, **3**, 285–291.
- Ihaka, R. and Gentleman, R. (1996) A language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.
- Irizarry, R.A. *et al.* (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
- Kropf, S. (2000) *Hochdimensionale Multivariate Verfahren in der Medizinischen Statistik*, Shaker Verlag, Aachen.
- Kropf, S. and Läuter, J. (2002) Multiple tests for different sets of variables using a data-driven ordering of hypotheses, with an application to gene expression data. *Biomet. J.*, **44**, 789–800.
- Läuter, J. *et al.* (1996) New multivariate tests for data with an inherent structure. *Biomet. J.*, **38**, 5–23.
- Läuter, J. *et al.* (1998) Multivariate Tests Based on Left-Spherically Distributed Linear Scores. *Ann. Stat.*, **26**, 1972–1988.
- Li, C. and Wong, W.H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA*, **98**, 31–36.
- Reiner, A. *et al.* (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, **19**, 368–375.
- Schuster, E. *et al.* (2004) Microarray based gene expression analysis using parametric multivariate tests per gene—a generalized application of multiple procedures with data-driven order of hypotheses. *Biomet. J.*, **46**, 687–696.
- Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
- Westfall, P.H. *et al.* (2004) Weighted FWE-controlling methods in high-dimensional situations. In Benjamini, Y., Bretz, F. and Sarkar, S.K. (eds), *Recent Developments in Multiple Comparison Procedures*, IML Lecture Notes and Monograph series, pp. 143–154.
- Westfall, P.H. and Young, S.S. (1993) *Resampling-Based Multiple Testing: Examples and Methods for Multiple P-Value Adjustment*. John Wiley & Sons, New York.