

Effects of threshold choice on the results of gene expression profiling, using microarray analysis, in a model feeding experiment with rats

ANJA HARTMANN¹, GERD NUERNBERG², DIRK REPSILBER², PAWEL JANCZYK⁵, CHRISTINA WALZ¹, SIRILUCK PONSUKSILI¹, WOLFGANG-BERNHARD SOUFFRANT³ and MANFRED SCHWERIN^{1,4}

¹Research Group Functional Genomics, ²Research Unit Genetics and Biometrics, ³Research Unit Nutritional Physiology »Oskar Kellner«, Research Institute for the Biology of Farm Animals (FBN), Dummerstorf, Germany, ⁴Institute of Farm Animal Sciences and Technology, University of Rostock, Rostock, Germany, ⁵Institute of Veterinary Anatomy, Department of Veterinary Medicine, Free University of Berlin, Berlin, Germany

Abstract

Global gene expression studies using microarray technology are widely employed to identify biological processes which are influenced by a treatment e.g. a specific diet. Affected processes are characterized by a significant enrichment of differentially expressed genes (functional annotation analysis). However, different choices of statistical thresholds to select candidates for differential expression will alter the resulting candidates list. This study was conducted to investigate the effect of applying a False Discovery Rate (FDR) correction and different fold change thresholds in statistical analysis of microarray data on diet-affected biological processes based on a significantly increased proportion of differentially expressed genes. In a model feeding experiment with rats fed genetically modified food additives, animals received a supplement of either lyophilized inactivated recombinant VP60 baculovirus (rBV-VP60) or lyophilized inactivated wild type baculovirus (wtBV). Comparative expression profiling was done in spleen, liver and small intestine mucosa. We demonstrated the extent to which threshold choice can affect the biological processes identified as significantly regulated and thus the conclusion drawn from the microarray data. In our study, the combined application of a moderate fold change threshold ($FC \geq 1.5$) and a stringent FDR threshold ($q \leq 0.05$) exhibited high reliability of biological processes identified as differentially regulated. The application of a stringent FDR threshold of $q \leq 0.05$ seems to be an essential prerequisite to reduce considerably the number of false positives. Microarray results of selected differentially expressed molecules were validated successfully by using real-time RT-PCR.

Keywords: False Discovery Rate, fold change threshold, microarray expression profiling, biological process

Zusammenfassung

Der Einfluss unterschiedlicher Schwellenwerte auf die Ergebnisse von Expressionsanalysen mittels Microarray Technik in einem Modell-Fütterungsversuch mit Ratten

Globale Genexpressionsanalysen mittels Mikroarray Technologie werden vielfach zur Identifizierung biologischer Prozesse, die durch eine Behandlung wie z. B. eine bestimmte Diät beeinflusst sind, eingesetzt. Betroffene Prozesse sind durch eine signifikant erhöhte Anzahl von different exprimierten Genen (»functional annotation analysis«) gekennzeichnet. Allerdings ändert sich bei Anwendung unterschiedlicher Grenzwert-Parameter zur Bestimmung von Kandidatengenen für differente Expression die daraus resultierende Kandidaten-Liste. Diese Untersuchungen wurden durchgeführt, um den Einfluss der Anwendung einer »False Discovery Rate« (FDR)-Korrektur und von Verhältnis-(FC)-Schwellenwerten in der statistischen Analyse von Microarray-Daten auf die Identifizierung von Diät-beeinflussten biologischen Prozessen, basierend auf einem signifikant erhöhten Anteil different exprimierter Gene, zu untersuchen. In einem Modellfütterungsexperiment mit Ratten, denen ein genetisch veränderter Futtermittelzusatz verabreicht wurde, erhielten die Tiere als Zusatz entweder lyophilisierte, inaktivierte rekombinante VP60-Bakuloviren (rBV-VP60) oder lyophilisierte inaktivierte Wildtyp-Bakuloviren (wtBV). Ein vergleichendes Expressionsprofiling wurde in Milz, Leber und in der Dünndarm-Mucosa durchgeführt. Unsere Untersuchungen verdeutlichen, in welchem Ausmaß die Wahl des Grenzwertes die Identifizierung signifikant beeinflusster biologischer Prozesse und damit auch die Schlussfolgerungen, die aus den Mikroarray Daten gezogen werden, beeinflussen kann. Es konnte gezeigt werden, dass die Kombination von moderaten FC- ($FC \geq 1,5$) und stringenten FDR-Schwellenwerten ($q \leq 0.05$) zur verlässlichen Identifikation biologischer Prozesse führte, die als different reguliert identifiziert wurden. Die Anwendung eines stringenten FDR-Schwellenwertes von $q \leq 0.05$ scheint notwendig, um die Anzahl der »Falsch-Positiven« bedeutsam zu reduzieren. Die Expressionswerte einiger ausgewählter, different exprimierter Moleküle der Mikroarray Ergebnisse konnten erfolgreich mittels Echtzeit-RT-PCR bestätigt werden.

Schlüsselwörter: False Discovery Rate, Ratio Grenzwert, Mikroarray Expressionsprofiling, biologischer Prozess

Introduction

DNA chip (microarray) technology is a modern and powerful method for performing global gene expression analyses and allows a characterization of changes in the multi-gene expression pattern, providing clues about regulatory mechanisms, cellular functions and biochemical pathways (LOCKHART & WINZELER 2000). In correspondence, the examination of large scale gene expression changes could be a sensitive and useful method for physiological characterization of animals, like e.g. for studying physiological side effects of genetically modified (GM) foods (LIU-STRATTON *et al.* 2004) in the light of known impacts of macro- and micronutrients as well as dietary bioactive molecules on gene expression (DE CATERINA & MADONNA 2004). Consequently, DNA chip technology could contribute to an increased probability of detecting unexpected physiological side effects due to genetically modified food or food components.

However, methods that are used for data analysis of microarray experiments can have a profound influence on interpretation of the results (QUACKENBUSH 2001). For the determination of genes that show significant expression changes between experimental groups due to a diet, a t-test is usually carried out. According to literature p -values without multiplicity adjustment or False Discovery Rate (FDR) corrected p -values with different stringencies (e.g. from 0.05-0.2) are used in combination with different fold changes (≥ 1.3 to 2.0) as cut-off thresholds to identify genes, which are »differentially expressed« between the corresponding experimental groups (BOOTH *et al.* 2004, RIEGER and CHU 2004, RUIZ-BALLESTEROS *et al.* 2005, ZHU *et al.* 2005, HAN *et al.* 2006, CHOI *et al.* 2007, INGRAM *et al.* 2007).

It is comprehensible that using different thresholds and filter criteria can have a profound influence on the list of differentially expressed genes (DEGs) (PAN *et al.* 2005). These lists of DEGs in turn are often the base for the identification of biological processes and thus for the conclusion to be drawn from the results.

In a model feeding experiment physiological effects of a transgenic food additive, consisting of lyophilized inactivated recombinant VP60 baculovirus (rBV-VP60), was investigated in rats.

We intended to study the effect of FDR correction using different threshold stringencies (without FDR: $p \leq 0.05$; with FDR: $q \leq 0.2$, $q \leq 0.1$, $q \leq 0.05$) and of fold change filters (without FC, $FC \geq 1.5$, $FC \geq 2.0$) on the biological processes identified as being diet-affected. Expression data of selected genes were verified by quantification of transcript levels using real-time RT-PCR.

Material and methods

Animals, diets and experimental design

For the feeding trial 13 young male Wistar rats (Charles River Laboratories, Germany) were used. The animals were housed in groups for 42 days. For the experiment the rats were divided into two groups (6 rats in the control group, 7 rats in the test group). A commercial basic diet (ALTROMIN Standard-Diet 1310) supplemented with 15% of lyophilized potatoes was fed to both groups. In addition, both groups received as fodder additive 20 ml of a solution consisting of corresponding amounts (approximately $10^{6.0}$ TCID₅₀ per ml) of lyophilized and ethylenimin-inactivated baculovirus. The control group was fed a non-transgenic wildtype baculovirus (wtBV) and the test group was given the transgenic baculovirus (rBV-VP60: $\sim 1 \mu\text{g}$ VP60/ml; VP60 is a major structural protein of a calicivirus that causes the rabbit hemorrhagic disease) which was propagated and handled according to HAMMER (2006). The baculovirus additives were fed on days 0, 2, 4 and 21, 23, 25. Both diets were isoenergetic and isonitrogenous. Animals had free access to the food and water.

At the end of the experiments animals were slaughtered and tissue samples of spleen, liver and small intestine mucosa were collected, immediately frozen in liquid nitrogen and stored at -80°C until isolation of RNA.

RNA extraction

Comparative expression analyses of the rats fed food additives with rBV-VP60 and wtBV, respectively, were done in spleen, small intestine mucosa and liver. After homogenization in Trizol reagent the total RNA was extracted from tissue samples using RNeasy Easy Kit (QIAGEN) according to manufacturers' instructions. Expression analyses were carried out using representative group specific RNA pools of liver, spleen and small intestine mucosa. RNA pools of each tissue and group were made up of equal amounts of RNA aliquots of the individual undiluted RNA samples.

Microarray analysis

For genome wide expression profiling we used the rat 10K oligo-array (MWG Biotech AG), which allows the detection of expression levels of 9802 transcripts. A dual labeling system was applied (cy3 and cy5 dyes). Probe preparation and chip hybridization were carried out essentially according to manufacturers' protocol (MWG Biotech AG). Each hybridization was carried out in triplicate.

All chips were scanned with a 428TM Scanner (Affymetrix) at 532 nm (cy3) and 635 nm (cy5). Fluorescence intensity of each spot was measured and a quality check was carried out using ImaGene5 software (BioDiscovery Inc.). Each signal had to pass a number of quality criteria including consistency of signal, minimum intensity level (cy3-signal or cy5-signal ≥ 150) and minimum signal-to-background ratio ($[\text{mean signal of spot} - \text{mean signal of background}] / \text{standard deviation [SD] of background signal} \geq 2$). Intensity values for each chip were normalized using the Lowess routine of GeneSight 3.0 (BioDiscovery Inc.). Genes that passed these criteria were included in further statistical analysis.

Table 1
Primers used for real-time RT-PCR amplification (40 cycles)

Primer für die Echtzeit-RT-PCR Amplifikation

Locus name ^a	Acc. no.	Primer sequences (5'-3')		AT ^b (°C)	Length PCR prod. (bp)	FT ^c (°C)
		sense primer	anti-sense primer			
Pla2g1b	NM_031585	TCCTCATCGACAACCCCTAC	ACGGCATAGACAGGAAGTGG	60	217	86
Pnlip	NM_013161	AACCTTACTTCCAGGGCACA	GTCAACGATTTGGGAAAAGGA	60	206	83
Cebpb	NM_024125	CAAGCTGAGCGACGAGTACA	CAGCTGCTCCACCTTCTTCT	63	157	90
Rara	AJ002940	CCCTACGCCCTTCTTCTTCC	AGGGCTGGGCACTATCTCTT	60	159	87
Gapdh	NM_017008	CATGGCCTTCCGTGTCCTA	CCCAGCATCAAAGGTGGAAG	65	205	86
Ctrb1	NM_012536	CAAATACAATGCCCTCAAGACC	CAGGAGTGGAAAGTGGAAACAGA	65	245	89
Gp5	NM_012795	AGGGACCTTGACAAAACCT	CGCGATAAATCCAACACCTT	63	134	81
F3	NM_013057	GCTGGGAACCTGAGATTG	CCTGGGCAACCTAGTGTCTT	63	200	83
Add2	NM_012491	GCTTCTCCATGAACCTTCTCC	GTCCTGCTCCTTGCTCACTC	60	189	87
Ela1	NM_012552	CATCTGCTCCAGCTCCTCTT	GTGAAGACTGTGGCTTCTT	60	204	87

^aPla2g1b phospholipase A2, group IB, Pnlip pancreatic lipase, Cebpb CCAAT/enhancer binding protein, beta, Rara retinoic acid receptor, alpha, Gapdh glyceraldehyde-3-phosphate dehydrogenase, Ctrb1 chymotrypsinogen B1, Gp5 glycoprotein V (platelet), F3 coagulation factor III (thromboplastin, tissue factor), Add2 adducin 2 (beta), Ela1 elastase 1, pancreatic, ^bannealing temperature, ^cfluorescence acquisition temperature

Real-time RT-PCR

Synthesis of first strand cDNA was performed with MMLV-RT (Promega, Madison, USA) and poly-T primers using 2 µg pooled total RNA of the corresponding feeding group. Quantitative analysis of PCR products was carried out in the LightCycler instrument (Roche, Mannheim, Germany) according to optimized PCR protocols as described by SCHWERIN *et al.* (2006). Gene specific primers, annealing temperature and fluorescence acquisition temperature used are given in Table 1. For all assays a standard curve was generated using DNA standard dilutions (10^2 , 10^3 , 10^4 , 10^5 and 10^6 copies) of the corresponding PCR product. The copy number of the housekeeping gene Gapdh was measured to normalize for equal RNA amounts. The mRNA abundance was analyzed in triplicate.

Statistical analysis

Single gene analysis: For microarray studies the expression signals of RNA pools of the control and the test group for each tissue were compared. Fold changes were calculated for each gene (signal intensity of rBV-VP60/signal intensity of wtBV). Significance of differential expression was analyzed by t-test of SAS (2002). To correct for multiple testing and to control the False Discovery Rate (FDR) we used the concept of STOREY and TIBSHIRANI (2003) and the software *Q-value* which calculates a corresponding *q-value* for the ordered *p-values*. Several FDR thresholds (No FDR: $p \leq 0.05$, FDR: $q \leq 0.05/0.1/0.2$) and fold change filters (no FC, $FC \geq 1.5/2.0$) were tested to identify differentially expressed genes and biological processes, respectively.

Functional annotation analysis: Based on Gene Ontology (GO) annotation (<http://www.geneontology.org>) biological processes with a significantly increased number of differentially expressed genes were arrogated to have an Ease Score ≤ 0.05 . The Ease Score is implemented in the DAVID software which calculates the enrichment with respect to the total number of genes assayed and annotated within the classification system. The Ease Score is a conservative adjustment of the fisher exact probability, taking into account the overestimation of categories assigned with only few genes (HOSACK *et al.* 2003). Not for all genes a functional annotation was available. In spleen 3738, in small intestine mucosa 3600 and in liver 4022 genes were GO annotated. Thus, for the identification of biological processes only these molecules could be included.

Real-time RT-PCR: Transcript levels of selected molecules were determined using real-time RT-PCR and the two experimental groups were compared with t-test of SAS (2002). Significance was assumed at $p \leq 0.05$.

Results

Effect of a FDR correction on the number of identified differentially expressed genes and on biological processes with a significant enrichment of DEGs in liver, spleen and small intestine mucosa

We studied the influence of applying different thresholds, on the number of genes identified as differentially expressed and on biological processes identified as being diet-affected. Transcript levels were determined in spleen, small intestine mucosa and liver of

rats fed either rBV-VP60, a genetically modified (GM) food additive or wtBV, a non-GM additive. Of 9802 total probes on chip 8800, 8339 and 9320 in spleen, small intestine and liver, respectively (see Table 2) passed the quality criteria and could be included in further analyses. Expression levels of both feeding groups were compared.

Firstly, we tested the impact of an application of a FDR correction using different significance thresholds ($q \leq 0.05$, $q \leq 0.1$, $q \leq 0.2$) on the quantity of genes identified as differentially expressed between both feeding groups. The number of differentially expressed genes was comparable in all three tissues (spleen, 2377, liver, 2754, small intestine mucosa, 1608 genes) when applying a t-test ($p \leq 0.05$) without FDR correction. However, the number of expected false positives was very high (spleen: 440 genes, small intestine mucosa: 417 genes, liver: 466 genes). In contrast, when applying a t-test plus FDR correction and a commonly used threshold ($q \leq 0.05$), the number of DEGs was reduced quite differently in the three tissues. Whereas, in the liver and the spleen 95% and 39% of the genes, respectively, were still identified as differentially expressed, in intestine less than 1% of the genes were left as differentially expressed, indicating that most if not all DEGs identified in intestine only by t-test were false positive ones. When using FDR correction with lower stringency ($q \leq 0.1$ and 0.2), the number of DEGs increased. At these FDR rates the quantity of DEGs was even higher than after t-test without FDR correction. However, along with more DEGs, an increased amount of false positives has to be expected.

Table 2

Number of differentially expressed genes identified at different »False Discovery Rate« (FDR) thresholds in spleen, liver and small intestine mucosa of rats fed rBV-VP60 in comparison with rats fed wtBV

Anzahl der mit unterschiedlichen »False Discovery Rate«-(FDR)-Schwellenwerten identifizierten, different exprimierten Gene in der Milz, der Leber und der Dünndarm-Mucosa von Ratten. Die differente Expression bezieht sich auf den Vergleich zwischen rBV-VP60 gefütterten und wtBV gefütterten Tieren.

Tissue	T-test with or without FDR correction	Differentially expressed genes	Expected »false positive« genes
Spleen ($n=8800$ analyzed genes)	No FDR ($p \leq 0.05$)	2377	440
	FDR ($q \leq 0.2$)	6241	1248
	FDR ($q \leq 0.1$)	3667	367
	FDR ($q \leq 0.05$)	917	46
Liver ($n=9320$ analyzed genes)	No FDR ($p \leq 0.05$)	2754	466
	FDR ($q \leq 0.2$)	7759	1552
	FDR ($q \leq 0.1$)	5241	524
	FDR ($q \leq 0.05$)	2620	131
Small intestine mucosa ($n=8339$ analyzed genes)	No FDR ($p \leq 0.05$)	1608	417
	FDR ($q \leq 0.2$)	4560	912
	FDR ($q \leq 0.1$)	355	35
	FDR ($q \leq 0.05$)	10	1

Based on the differentially expressed genes recognized in liver and spleen, diet-affected biological processes (GO annotation) with a significant enrichment (Ease Score ≤ 0.05) of DEGs were identified by means of DAVID software. In Table 3a and 3b biological processes of spleen and liver with a significant over-representation of DEGs are shown in dependence on the FDR threshold that was applied.

Table 3

Influence of applying different »False Discovery Rate« (FDR) thresholds on biological processes exhibiting a significantly increased number of differentially expressed genes. Different gene expression is referring to spleen (a) and liver (b) of rats fed rBV-VP60 in comparison with rats fed wtBV.

Einfluss der Anwendung verschiedener »False Discovery Rate«-(FDR)-Grenzwerte auf die Identifizierung von biologischen Prozessen, die eine signifikante Anreicherung von different exprimierten Genen zeigen. Die differente Expression bezieht sich auf den Vergleich zwischen rBV-VP60 und wtBV gefütterten Ratten in Milz und Leber.

3a – Spleen

Biological process (GO term)	GO annotated genes	Differentially expressed GO annotated genes							
		without FDR		FDR		FDR		FDR	
		$p \leq 0.05$		$q \leq 0.2$		$q \leq 0.1$		$q \leq 0.05$	
	n^1	ES^2	n^1	ES^2	n^1	ES^2	n^1	ES^2	
nucleic acid metabolism	719	216	0.03	513	0.05	326	0.01	58	0.08
central nervous system development	103	38	0.03	71	0.60	46	0.34	7	0.69
aldehyde metabolism	10	7	0.03	8	0.61	8	0.06	2	0.50
hyperosmotic response	6	5	0.05	5	0.72	5	0.20	3	0.06
DNA repair	51	19	0.12	44	0.01	25	0.24	5	0.44
establishment of cellular localization	269	73	0.55	201	0.02	124	0.07	19	0.53
cellular localization	274	74	0.57	204	0.02	125	0.09	20	0.46
intracellular transport	265	72	0.54	197	0.03	122	0.08	19	0.50
secretory pathway	129	35	0.59	99	0.04	57	0.34	8	0.77
cell migration	146	39	0.63	111	0.04	64	0.35	15	0.10
organelle organization and biogenesis	267	77	0.30	197	0.04	121	0.12	18	0.62
regulation of protein metabolism	108	31	0.46	70	0.88	56	0.03	12	0.10
biopolymer metabolism	673	192	0.18	461	0.56	300	0.04	44	0.65
macromolecule catabolism	119	36	0.30	87	0.21	60	0.04	10	0.40
response to salt stress	4	3	0.29	3	0.90	3	0.55	3	0.02
lipid catabolism	54	16	0.49	34	0.91	19	0.91	9	0.02
cellular lipid metabolism	265	67	0.80	187	0.30	102	0.88	27	0.03
glycerol ether metabolism	12	6	0.20	11	0.22	7	0.37	4	0.04
<i>all</i>	<i>3738</i>	<i>1007</i>		<i>2562</i>		<i>1544</i>		<i>250</i>	

3b – Liver

Biological process (GO term)	GO annotated genes	Differentially expressed GO annotated genes							
		without FDR		FDR		FDR		FDR	
		$p \leq 0.05$		$q \leq 0.2$		$q \leq 0.1$		$q \leq 0.05$	
	n^1	ES^2	n^1	ES^2	n^1	ES^2	n^1	ES^2	
nucleic acid metabolism	760	248	0.05	628	0.01	441	0.08	239	0.04
regulation of Wnt receptor signaling pathway	14	9	0.03	12	0.67	10	0.35	9	0.02
macromolecule catabolism	124	47	0.05	99	0.56	75	0.20	45	0.06
spermatid development	16	11	0.01	14	0.56	12	0.21	9	0.06
aldehyde metabolism	11	8	0.02	10	0.59	10	0.07	7	0.06
germ cell development	19	12	0.01	18	0.21	17	0.01	10	0.06
epidermis morphogenesis	13	7	0.16	12	0.47	12	0.03	6	0.31
transport	1120	339	0.44	915	0.01	650	0.03	329	0.30
organelle organization and biogenesis	275	86	0.38	237	0.00	160	0.23	80	0.51
DNA repair	50	17	0.43	46	0.03	30	0.42	16	0.47
sensory perception of smell	50	13	0.86	46	0.03	30	0.42	13	0.81
<i>all</i>	<i>4022</i>	<i>1205</i>		<i>3183</i>		<i>2234</i>		<i>1153</i>	

¹ n number of differentially expressed genes in process, ² ES calculated Ease Score based on a Fisher's exact probability, significance is assigned at $ES \leq 0.05$, grey colour marks significantly affected biological processes

An application of a FDR threshold considerably modified type and number of significantly enriched biological processes. Table 3 shows clearly that the decision to accept higher rates of false positives (10% and 20%) can change the results dramatically. Therefore, together with FDR thresholds, fold change filters were additionally applied based on the assumption that among genes showing low expression differences the probability of false positive genes is higher. Small intestine was excluded from the further study because only few genes remained differentially expressed after FDR correction.

Effect of a combined application of FDR correction based thresholds and fold change filters on the number differentially expressed genes and on biological processes with a significant enrichment of DEGs in liver and spleen

We tested a combined application of different FC filters and FDR thresholds on the number of DEGs. Moreover we examined the impact of FC filters on the pattern of significantly enriched biological processes when a stringent FDR threshold ($q \leq 0.05$), which is commonly used, was applied.

Table 4

Number of differentially expressed genes, identified with different »False Discovery Rate« (FDR) and »fold change« (FC) thresholds in spleen and liver of rats fed rBV-VP60 in comparison with rats fed wtBV.

Anzahl der mit unterschiedlichen »False Discovery Rate«-(FDR)- und »Verhältnis«-(FC)-Schwellenwerten identifizierten, different exprimierten Gene in der Milz und Leber von Ratten. Different Expression bezieht sich auf den Vergleich zwischen rBV-VP60 gefütterten und wtBV gefütterten Tieren.

Tissue	T-test with or without FDR correction	Number of differentially expressed genes		
		No FC	FC ≥ 1.5	FC ≥ 2.0
Spleen (n=8800 analyzed genes)	No FDR ($p \leq 0.05$)	2377	248	66
	FDR ($q \leq 0.2$)	6241	279	73
	FDR ($q \leq 0.1$)	3667	268	68
	FDR ($q \leq 0.05$)	917	157	49
Liver (n=9320 analyzed genes)	No FDR ($p \leq 0.05$)	2754	153	24
	FDR ($q \leq 0.2$)	7759	242	42
	FDR ($q \leq 0.1$)	5241	230	39
	FDR ($q \leq 0.05$)	2620	149	24

In Table 4 the number of differentially expressed genes identified by using different thresholds is shown for the tissues spleen and liver. When applying FC filters of ≥ 1.5 or ≥ 2.0 the number of DEGs was reduced considerably at all FDR rates and in both tissues. Without FDR correction the application of a $FC \geq 1.5$ lead to a reduction of DEGs from 2377 to 248 in spleen and from 2754 to 153 in liver, which constitute 10% and 6% of the DEGs found without FC filter. A FC filter of ≥ 2 lead to a further reduction of DEGs; only 3% and 1%, respectively remained significant in spleen and liver. With a FDR correction ($q \leq 0.05$) the influence of a FC filter of ≥ 1.5 was just as severe. The number of DEGs decreased from 917 to 157 genes (17%) in spleen and from 2620 to 149 genes (6%) in liver. When raising FC-filter stringency to ≥ 2 only 5% of the DEGs in spleen and 1% in liver remained.

The impact of a FC cut-off on the number of significantly diet-affected biological processes was likewise strong. After applying a FDR threshold with a more stringent $q \leq 0.05$ in combination with a moderate $FC \geq 1.5$ filter, 1 and 7 biological processes were identified as differentially regulated in spleen and liver, respectively (see Table 5a, b).

Table 5

Influence of applying different »fold change« filters on identified biological processes exhibiting a significant enrichment of differentially expressed genes. Different gene expression is referring to spleen (a) and liver (b) of rats fed rBV-VP60 in comparison with rats fed wtBV and to DEGs with q -values ≤ 0.05 .

Einfluss der Anwendung verschiedener »Verhältnis«-Schwellenwerte auf die Identifizierung von biologischen Prozessen, die eine signifikante Anreicherung von differenzial exprimierten Genen zeigen. Differente Expression bezieht sich auf den Vergleich zwischen rBV-VP60 gefütterten und wtBV gefütterten Tieren in den Geweben Milz und Leber und auf Gene mit q -Werten ≤ 0.05 .

5a – Spleen

Biological process (GO term)	GO annotated genes in process	Differentially expressed GO annotated genes					
		no FC		FC ≥ 1.5		FC ≥ 2	
		n^1	ES ²	n^1	ES ²	n^1	ES ²
lipid catabolism	54	9	0.02	5	0.03	4	0.01
cellular lipid metabolism	265	27	0.03	6	0.72	3	0.54
glycerol ether metabolism	12	4	0.04	2	0.24	2	0.08
response to salt stress	4	3	0.02	1	1.00	–	–
<i>all</i>	3738	250		85		26	

5b – Liver

Biological process (GO term)	GO annotated genes in process	Differentially expressed GO annotated genes					
		no FC		FC ≥ 1.5		FC ≥ 2	
		n^1	ES ²	n^1	ES ²	n^1	ES ²
regulation of Wnt receptor signaling pathway	14	9	0.02	–	–	–	–
nucleic acid metabolism	760	239	0.04	25	0.01	–	–
metanephros development	18	6	0.62	4	0.01	1	1.00
positive regulation of metabolism	173	51	0.50	10	0.01	–	–
regulation of cellular metabolism	566	174	0.15	21	0.01	1	1.00
kidney development	22	6	0.80	4	0.01	1	1.00
intracellular signaling cascade	495	142	0.55	17	0.04	2	0.82
alcohol metabolism	164	54	0.16	8	0.05	1	1.00
<i>all</i>	4022	1153		83		14	

¹ n number of differentially expressed genes in process ²ES calculated Ease Score based on a Fisher's exact probability, significance is assigned at ES ≤ 0.05 , grey colour marks significantly affected biological processes

In liver most differentially expressed genes were involved in »nucleic acid metabolism«, »intracellular signaling cascade« and »regulation of cellular metabolism«, whereas in spleen they belonged to »lipid metabolism«. With a FC filter of ≥ 2 and q -value ≤ 0.05 only one process, namely »lipid catabolism« remained significant in spleen, whereas in liver no significant process was found due to the drastically shortened list of DEGs (compare Table 4).

Quantitative expression analysis using real-time RT-PCR

In order to confirm results obtained from the microarray study we performed real-time RT-PCR for selected molecules. Transcription levels of 9 genes (liver: 2, spleen: 7), which were identified as differentially expressed between both feeding groups in microarray analysis (FC ≥ 1.5 , $q \leq 0.5$), were re-analyzed by real-time RT-PCR. In Table 6 (spleen) and 7 (liver) relative transcript levels (mRNA abundance of rBV-VP60 fed rats/mRNA abundance of wtBV fed rats) of these genes measured by microarray analysis are compared with those measured by real-time RT-PCR. For 7 genes (*Pnlip*, *Pla2gl*, *Ctrb*, *Ela1*, *Add2*, *Rara*,

Cebpb) significant different transcript levels were observed between both feeding groups after real-time RT-PCR confirming results of microarray analysis. For the genes *Gp5* and *F3* microarray results could not be validated. Transcript levels measured by real-time RT-PCR between both feeding groups were not significantly different for these two candidates.

Table 6

Relative transcript levels (mRNA abundance of rBV-VP60 fed animals/mRNA abundance of wtBV fed animals) of selected genes measured in the spleen of rats with microarray technology and real-time RT-PCR.

Relative Transkriptniveaus (mRNA-Menge der rBV-VP60 gefütterten Tiere/mRNA-Menge der wtBV gefütterten Tiere) ausgewählter Gene in der Milz von Ratten nach Messung mit Microarray Technologie und Echtzeit-RT-PCR.

Gene name	Gene symbol	Microarray analysis		Real-time RT-PCR	
		RTL	q ¹	RTL	p ²
<i>pancreatic lipase</i>	<i>Pnlip</i>	0.47	0.03	0.64	0.00
<i>phospholipase A2, group IB</i>	<i>Pla2g1b</i>	0.47	0.05	0.52	0.00
<i>chymotrypsino-gen B</i>	<i>Ptrb</i>	0.63	0.05	0.56	0.00
<i>elastase 1, pancreatic</i>	<i>Ela1</i>	0.48	0.00	0.38	0.00
<i>adducin 2, beta</i>	<i>Add2</i>	1.71	0.03	1.45	0.01
<i>glycoprotein 5</i>	<i>Gp5</i>	0.52	0.05	1.15	0.43
<i>coagulation factor III</i>	<i>F3</i>	0.65	0.03	1.07	0.59

¹FDR corrected *p*-value of t-test, ²*p*-value of t-test, RTL relative transcript level

Table 7

Relative transcript levels (mRNA abundance of rBV-VP60 fed animals /mRNA abundance of wtBV fed animals) of selected genes measured in the liver of rats with microarray technology and real-time RT-PCR.

Relative Transkriptniveaus (mRNA-Menge der rBV-VP60 gefütterten Tiere/mRNA-Menge der wtBV gefütterten Tiere) ausgewählter Gene in der Leber von Ratten nach Messung mit Microarray Technologie und Echtzeit-RT-PCR.

Gene name	Gene symbol	Microarray		Real-time RT-PCR	
		RTL	q ¹	RTL	p ²
<i>retinoic acid receptor, alpha</i>	<i>Rara</i>	0.61	0.04	0.51	0.00
<i>CCAAT/enhancer binding protein (C/EBP), beta</i>	<i>Cebpb</i>	0.60	0.03	0.50	0.01

¹FDR corrected *p*-value of t-test, ²*p*-value of t-test, RTL relative transcript level

Discussion

The present study aimed to investigate the effect of applying FDR correction thresholds and FC filters with different stringencies to a microarray dataset used for functional annotation analysis, as a prerequisite for a reliable data interpretation. In a model feeding experiment with rats fed either lyophilized inactivated recombinant VP60 baculovirus (rBV-VP60) or lyophilized inactivated wild type baculovirus (wtBV) food additives, we used DNA chip technology to identify significantly affected biological processes. Based on the genes recognized as differentially expressed and on gene ontology (GO) annotation (<http://www.geneontology.org>), biological processes with a significantly increased number of differentially expressed genes were identified using a modified Fisher exact test (Ease Score ≤ 0.05), which is implemented in the DAVID software used.

To identify treatment related differences in gene expression, t-statistic and fold change based filtering is commonly used. Besides, it is recommended to correct for high numbers of false positives, which result from thousands of simultaneous t-tests (multiple

testing problem) in a typical microarray experiment (ALLISON *et al.* 2006). A FDR based approach can be applied to adjust for this problem. FC cut-offs with different stringencies are widely used in combination with t-statistics to define relevant expression changes. However, the question which fold change or which FDR threshold should be used remains unanswered from literature. Fold change thresholds of ≥ 2 and a FDR correction with $q \leq 0.05$ are common. Nevertheless, fold change filters of 1.3 or 1.5 and FDR thresholds of $q \leq 0.1$ and $q \leq 0.2$ can be found in literature (BOOTH *et al.* 2004, RIEGER & CHU 2004, RUIZ-BALLESTEROS *et al.* 2005, ZHU *et al.* 2005, HAN *et al.* 2006, CHOI *et al.* 2007, INGRAM *et al.* 2007).

By means of the model feeding experiment with rats gene expression changes, possibly caused by the GM diet, were analyzed in spleen, small intestine mucosa and liver. The number of genes identified as differentially expressed without using any thresholds except $p \leq 0.05$ was comparable in all three tissues analyzed (spleen: 2377, small intestine mucosa: 1608, liver: 2754). Applying an FDR threshold with a widely used level of significance ($q \leq 0.05$) to our dataset reduced the number of DEGs quite unequally in the three tissues. Whereas in liver a considerable number (95%) of the genes were still identified as differentially expressed, in spleen only 39% and in small intestine only 1% of the genes exceeded the threshold indicating a very high proportion of false positives in small intestine and partly in spleen. Lower stringencies (FDR $q \leq 0.1$ or 0.2) increased the number of DEG candidates, even exceeding the number of genes identified as differentially expressed after t-test *without* FDR correction. However, different inconsistent biological processes identified as significantly regulated support the assumption of a high proportion of false positives among the DEGs when using no or low stringent FDR conditions. The criteria for determining which FDR level to use as threshold should surely depend on the purpose and context of the experiment: If the purpose is to find as many DEGs as possible (maximizing sensitivity) higher rates of expected false positives are probably acceptable. If the objective is to identify a small number of truly differentially expressed genes (maximizing specificity) for further studies, a stringent threshold may be appropriate (PAWITAN *et al.* 2005, CHEN *et al.* 2007).

For our data we concluded that a FDR threshold of 5% would be most appropriate because a high reliability and a small amount of expected false positives were aimed. According to CHOE *et al.* (2005) many false positive candidates may occur at less stringent FDR thresholds. This may happen, because *basic t-statistics* are prone to false positives resulting from artificially low standard deviations, especially when only limited numbers of replicates are available, as it was the case in the present experiment. For experiments with large number of replicates variances may be estimated more correctly. In this case it might not be essential to apply a very stringent FDR of 5%. The MAQC Consortium (2006) has likewise recognized the unstable nature of the variance estimate in the t-statistic measure in the context of their investigations. Consequently they proposed the combined application of test statistics plus a fold change ranking for the identification of DEGs.

In consequence, we additionally applied fold change cut-offs to our data, based on the assumption that among genes showing low expression differences the probability of false positive genes is higher. The combined application of FDR (5%) and fold change thresholds ($FC \geq 1.5$, $FC \geq 2.0$) in data analysis did significantly reduce the numbers of

DEGs. In our study, a combination of FDR threshold (5%) with a moderate fold change cut-off of $FC \geq 1.5$ exhibited highest reliability regarding biological processes identified as differentially regulated. A FC of ≥ 2 was probably too stringent, because no significant biological process was identified in liver due to the low number of remaining DEGs with a $FC \geq 2$. In correspondence, it is recommended to apply smaller fold change cut-offs than ≥ 2 to involve not only excessive changes in fewer genes but also smaller adjustments of large gene sets (BEN-SHAUL *et al.* 2005). To ensure a high reliability of functional annotation analysis (identified biological processes by a significant enrichment of differentially expressed genes) we applied combined FDR and FC thresholds. According to our results, there is no alternative to an application of a stringent FDR threshold ($q \leq 0.05$) in order to reduce considerably the number of false positives. However, the decision, which fold change filter to use, should be taken in the context of the experiment to ensure a high stringency (further reduction of number of false positives) but also a critical number of remaining DEGs for the final Fisher's exact test. For this study a FC of ≥ 1.5 seems to fulfill these requests most appropriately.

Our results demonstrate the strong impact of threshold choice for defining DEGs on the results of functional annotation analysis. PAN *et al.* (2005) showed similar effects in their study. Their work was related to the identification of DEGs that comprise »signatures«, which are associated with tumour types and functional categories. By means of varying thresholds they produced »signatures« of different size and exhibited strong effects on the identification of functional categories. In correspondence to our study, it was concluded that microarray based gene signatures should be identified routinely by using a range of threshold conditions to assess the »sensitivity« of biological conclusions to threshold choice and the overall robustness of the conclusion.

As to be seen from our results the above discussed issues about threshold choice are not only a problem that scientists are facing when working in cancer or human related research. It is likewise of great importance in agricultural studies and, since microarray technology is more and more used in this field, the general need of such sensitivity analyses should be emphasized.

To summarize, it becomes clear that threshold choice considerably affects the fundamental outcome of microarray experiments. Probably, if thresholds shall be applied, they have to be chosen for each experiment individually and have to be taken with regard to the distinct experimental conditions and the objective of the study. In this context, a testing of a range of threshold conditions as proposed by PAN *et al.* (2005) should perhaps be considered as sensitivity analysis. However, suitable quality criteria are still missing, because we do not know the »truth«. That means that the decision, which thresholds are used to ensure a low number of false positives but also a critical number of remaining DEGs for the frequency statistics, will be always empirical. In contrast to GOEMAN *et al.* (2004), MANSMANN and MEISTER (2005) and NAM and KIM (2008) who discussed threshold-free functional annotation approaches (gene set enrichment) we propose the application of a stringent FDR threshold of $q \leq 0.05$ as an essential prerequisite to reduce significantly number of false positives. Although skipping a fold change cut-off probably increases the risk for high numbers of false positives (CHOE *et al.* 2005), as indicated by the large amount of DEGs that remained at all FDR levels in liver and spleen,

results of functional annotation analysis with or without the use of a FC filter showed essential agreement. In the liver and the spleen the most significant biological process identified using a conservative FDR threshold ($q \leq 0.05$) in combination with a FC of ≥ 1.5 (liver, »nucleic acid metabolism«; spleen, »lipid metabolism«) was also identified as significantly regulated using the FDR threshold without FC filter.

Acknowledgements

We thank Annette Jugert, Joana Bittner and Nicole Gentz for technical assistance. The used food additives (rBV-VP60 and wtBV) were kindly provided by Dr. Schirrmeier, Dr. Keil, Dr. König and Dr. Hammer. Financial support from the German Federal Ministry of Education, Science, Research and Technology (BMBF) (Project 0312744) is gratefully acknowledged.

References

- Allison DB, Cui X, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 7, 55-5
- Ben-Shaul Y, Bergman H, Soreq H (2005) Identifying subtle interrelated changes in functional gene categories using continuous measures of gene expression. *Bioinformatics* 21; 1129-37
- Booth S, Bowman C, Baumgartner R, Sorensen G, Robertson C, Coulthart M, Phillipson C, Somorjai RL (2004) Identification of central nervous system genes involved in the host response to the scrapie agent during preclinical and clinical infection. *JGenViro* 85, 3459-71
- Chen JJ, Wang SJ, Tsai CA, Lin CJ (2007) Selection of differentially expressed genes in microarray data analysis. *Pharmacogenomics J* 7, 212-20
- Choe SE, Boutros M, Michelson AM, Church GM, Halfon MS (2005) Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biol* 6, R16-R15
- Choi Y, Qin Y, Berger MF, Ballow DJ, Bulyk ML, Rajkovic A (2007) Microarray analyses of newborn mouse ovaries lacking nobox. *Biol Reprod* 77, 312-319
- De Caterina R, Madonna R (2004) Nutrients and gene expression. *World Rev Nutr Diet* 93, 99-133
- Goeman JJ, Van de Geer SA, de Kort F, van Houwelingen HC (2004) A global test for groups of genes: testing association with a clinical outcome *Bioinformatics* 20, 93-9
- Hammer M (2006) Expression of the capsid protein of the Rabbit Haemorrhagic Disease Virus in transgen plants: antigenicity and immunogenicity. PhD thesis, Tierärztliche Hochschule Hannover
- Han T, Wang J, Tong W, Moore MM, Fuscoe JC, Chen T (2006) Microarray analysis distinguishes differential gene expression patterns from large and small colony Thymidine kinase mutants of L5178Y mouse lymphoma cells. *BMC Bioinformatics* 7 Suppl 2 S9, 1-12
- Hosack DA, Dennis G JR, Sherman BT, Lane HC, Lempicki RA (2003) Identifying biological themes within lists of genes with EASE *Genome Biol* 4, R70-R75
- Ingram JL, Ntao-Menezes A, Turpin EA, Wallace DG, Mangum JB, Pluta LJ, Thomas RS, Bonner JC (2007) Genomic analysis of human lung fibroblasts exposed to vanadium pentoxide to identify candidate genes for occupational bronchitis. *Respir Res* 8:34
- Liu-Stratton Y, Roy S, Sen CK (2004) DNA microarray technology in nutraceutical and food safety. *Toxicol Lett* 150, 29-42
- Lockhart DJ, Winzeler EA (2000) Genomics gene expression and DNA arrays. *Nature* 405, 827-36
- Mansmann U, Meister R (2005) Testing differential gene expression in functional groups Goeman's global test versus an ANCOVA approach. *Methods Inf Med* 44, 449-53
- Maqc Consortium (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotech* 24, 1151-61
- Nam D, Kim SY (2008) Gene-set approach for expression pattern analysis Brief. *Bioinformatics* 9, 189-97
- Pan KH, Lih CJ, Cohen SN (2005) Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays. *Proc Natl Acad Sci USA* 102, 8961-5
- Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A (2005) False discovery rate sensitivity and sample size for microarray studies. *Bioinformatics* 21, 3017-24

- Quackenbush J (2001) Computational analysis of microarray data. *Nat Rev Genet* 2, 418-27
- Rieger KE, Chu G (2004) Portrait of transcriptional responses to ultraviolet and ionizing radiation in human cells. *Nucleic Acids Res* 32, 4786-03
- Ruiz-Ballesteros E, Mollejo M, Rodriguez A, Camacho FI, Algara P, Martinez N, Pollan M, Sanchez-Aguilera A, Menarguez J, Campo E, Martinez P, Mateo M, Piris MA (2005) Splenic marginal zone lymphoma: proposal of new diagnostic and prognostic markers identified after tissue and cDNA microarray analysis. *Blood* 106, 1831-8
- Schwerin M, Kuehn C, Wimmers S, Walz C, Goldammer T (2006) Trait-associated expressed hepatic and intestine genes in cattle of different metabolic type – putative functional candidates for nutrient utilization. *J Anim Breed Genet* 123, 307-14
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100, 9440-5
- Zhu D, Hero AO, Cheng H, Khanna R, Swaroop A (2005) Network constrained clustering for gene microarray data. *Bioinformatics* 21, 4014-20

Received 5 December 2008, accepted 6 January 2009.

Corresponding author:

Prof. Dr. MANFRED SCHWERIN

email: schwerin@fhn-dummerstorf.de

Research Institute for the Biology of Farm Animals, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany
