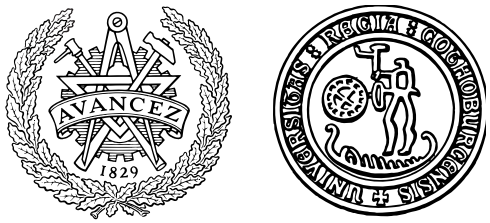
THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Contributions to Statistical Analysis of Gene Expression Data

Anders Sjögren

CHALMERS | GÖTEBORG UNIVERSITY



Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and Göteborg University
Göteborg, Sweden, 2005

Contributions to Statistical Analysis of Gene Expression Data
Anders Sjögren

Licentiate Thesis
ISSN 1652-9715/No. 2005:13

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31 772 1000

Printed in Göteborg, Sweden 2005

Contributions to Statistical Analysis of Gene Expression Data

Anders Sjögren

Abstract

The abundance of RNA copies (i.e. *gene expression*) for a certain gene in a cell determines the production of the corresponding protein, affecting the machinery of the cell. Using a technique called DNA microarrays, the gene expression of thousands of genes can be measured simultaneously, providing a snapshot of the activity of the measured cells. This thesis consists of three papers dealing with the statistical analysis of gene expression experiments.

In Paper I, an experiment is analysed aiming at identifying genes regulated by treatment of nasal polyps with local glucocorticoids . A cube root variance stabilising transformation is applied and a moderated t-statistic is computed, protecting against spurious significances. Finally, p-values are evaluated by forming a null distribution by a permutation procedure.

In Paper II, stably expressed genes are sought, to be used as references. Novel candidate reference genes are first identified using a collection of microarray datasets. The candidate genes and widely used ones are then examined closer, using a measurement technique demanding a reference gene to be known. A previously published technique providing a ranking of the genes is extended by a bootstrapping step, to assess the credibility of the ranks.

In Paper III, a novel model for the analysis of paired microarray experiments is proposed, taking precision differences between replicates (biological and/or technical) into account. The variance structure involves (i) gene-specific scaling factors with a prior distribution, moderating the highly variable variance estimates and (ii) a covariance matrix aiming at catching array-wide differences in quality, including shared sources of variation. Methods for parameter estimation are presented and a likelihood-ratio test is deduced. The procedure is compared to existing methods on both real and simulated data. On real data, substantial differences in quality between repetitions are found. On simulated data, improved performance is shown in some cases.

Acknowledgements

First of all, I would like to thank my advisors Mats Rudemo, Lena Carlsson and Olle Nerman for sharing their wisdom and intuition.

Furthermore, I would like to thank my PhD-student colleagues Erik Kristiansson and Louise Olofsson for the close and rewarding collaborations in the statistical and biological directions, respectively.

Margareta Jernås and Bob Olsson always have energetic and inspiring suggestions and ideas on biological matters, for which I am grateful.

I would also like to thank those of my fellow co-authors not mentioned above for the inspiring collaborations.

Without the financial support from The National Graduate School in Genomics and Bioinformatics, I would not be here.

Last, but certainly not least, I would like to thank my fiancée Camilla and the rest of my family for their unfailing love and support during the intense last few years. *Shared joy is a double joy; shared sorrow is half a sorrow.*

Background and introduction

In the world of molecular biology, the genome (DNA) is the passive blueprint describing how different proteins can be built. The proteins in turn are the actual working horses of the cells, determining the behaviour of the cell. In order to produce more of a certain protein, a working order is placed in the form of RNA copies of a relevant part of the DNA. Such a part of the DNA encoding a protein is called a gene. The fact that DNA is transcribed into RNA, which is translated into proteins is called the central dogma of molecular biology. RNA abundance is often referred to as gene expression.

During the last few years, a measurement technique called DNA microarrays has become widely used and substantially developed, both on the technical and the analytical side. One DNA microarray is able to measure the abundance of RNA for thousands of genes simultaneously, thus creating a snapshot of the activity of the measured cells.

In the current thesis, some contributions related to the statistical analysis of microarray experiments are presented. The variant of microarrays used are mainly *in situ* hybridised arrays, manufactured by Affymetrix Inc. (Santa Clara, USA). A good description of the steps involved in the low level analysis of Affymetrix type arrays is provided in Bolstad (2004). That includes several consecutive steps from the extracted RNA to the actual normalised array signals signifying the gene expression for the different genes.

One challenging problem in the statistical analysis of microarray data is that the datasets have thousands of dimensions (genes) but only relatively few (biological) repetitions, typically 3 to 10. The challenge is then to take advantage of this structure, using the fact that many genes behave similarly and that many genes are measured on the same array, potentially carrying shared characteristics. In this setting, using techniques such as the ordinary t-test to find differentially expressed genes is problematic, since the gene-specific variance estimates are highly variable due to the small number of replicates, giving rise to false positives caused by underestimates of the standard deviation. Methods have been proposed, using the information in all genes to moderate extreme estimates (Efron et al., 2001; Tusher et al., 2001). A variant of the method by Efron et al. (2001), is used in Paper I. In Paper III, empirical Bayes based methods (Baldi and Long, 2001; Lönnstedt and Speed, 2002; Smyth, 2004) are generalised to include quality control aspects, modelling the variability of the different repetitions (biological and/or technical). Here, the structure of the data is utilised to estimate the variance of each array, as well as correlations between the different arrays.

Comments on Paper I:
Gene profiling reveals increased expression of uteroglobin and other anti-inflammatory genes in glucocorticoid-treated nasal polyps.

In Paper I (Benson et al., 2004), human polyps from five patients are examined, before and after treatment with local glucocorticoids. The aim was to identify differentially expressed genes, thereby increasing our understanding about the mechanisms of the treatment. Biopsies were taken from the polyps, microarrays were run in duplicates from each biopsy and low-level analysis was performed to produce one signal value per gene and array.

The next step was to test all genes to identify which are differentially expressed. If an ordinary paired t-test ($|t_g| > c$ where $t_g = \sqrt{n}\frac{\bar{X}^g_\cdot}{s_g}$) would be applied on the averaged duplicates, a large number of genes with small differential expression would be called significant due to gravely underestimated standard deviation estimates, caused by large variability in the standard deviation estimate. However, it is possible to take advantage of the structure of the data, having 5 biological repetitions and over 20000 genes, to address the problem; The standard deviation estimates, s_g, can be moderated by adding a suitably chosen global penalising constant, s_0, thus forming the *penalized t-statistic* of Efron et al. (2001):

$$z_g = \sqrt{n}\frac{\bar{X}^g_\cdot}{s_g + s_0} \, .$$

However, if different groups of the genes have different variability, some groups will be favoured compared to others. When examining the data at hand, the standard deviation estimates shows a strong trend of increasing variability with increasing mean expression before and after treatment. Using a log-transform before creating X^g_i, a trend of decreasing variability with increasing expression was instead apparent. Therefore, a cube root transformation was selected, as described in Tusher et al. (2001), resulting in substantially smaller trends. The moderating constant, s_0, was chosen to the 90th percentile of s_g, according to Efron et al. (2001).

To determine the critical value, c_α, for a certain level, α, of the test,

$$|z_g| > c_\alpha \, ,$$

a permutation approach was used. Ideally, for genes unaffected by the treatment, all four arrays from the same individual would be exchangeable. Therefore, we can create pseudo-differences, Y^g_i, by first creating pseudo-averages,

averaging one array before and one after treatment, and then taking the difference between the pseudo-averages. The z_g based on Y_i^g,

$$z_g^0 = \frac{\sqrt{n}\bar{Y}_\cdot^g}{s_g^0 + s_0} ,$$

are distributed according to the null distribution of z_g, i.e. the distribution of z_g for genes without differential expression. This can be performed in four ways for each individual, resulting in $4^5 = 1024$ permutations. The null distribution of z_g was approximated by the combined empirical distribution of the 1024 sets of z_g^0. Comparing the distribution of z_g to the null distribution, the variability of z_g seems to be larger than predicted. Considering the experimental layout closer, the arrays are not exchangeable even for non-differentially expressed genes, since the replicates stem from the sample extracted RNA from the same sample. Therefore, systematic variation due to differences between polyps, timepoints or RNA extractions are not reproduced correctly. Therefore, the p-values reported were reported as approximate and a stringent limit was set for calling a gene differentially expressed ($p < 0.001$).

Briefly, the biological results of the paper were that treating polyps with local glucocorticoids increase the expression of uteroglobin and other anti-inflammatory genes. The increased expression of uteroglobin was verified on the RNA level using realtime RT-PCR and on protein level using immunohistochemistry.

Comments on Paper II:
Evaluation of reference genes for studies of gene expression in human adipose tissue.

In Paper II (Gabrielsson et al., 2005), a non-typical problem is addressed. In certain measurement techniques, such as realtime RT-PCR, the gene expression for different genes can be measured, scaled by an unknown constant. Genes that are known to be stably expressed between the studied conditions may then serve as references to determine the scale. Widely used reference genes, e.g. β-actin and GAPDH, have been shown to be regulated under certain conditions and are therefore debated. Hence, Paper II aims at identifying a suitable reference gene to be used in studies involving human adipose tissue (i.e. fat), examining new candidates and widely used reference genes.

First, novel reference gene candidates are sought using three microarray datasets, involving a variety of conditions for adipose tissue. Here, genes are sought that are estimated to vary as little as possible between the conditions.

In the second part, the stability is evaluated using realtime RT-PCR samples for the novel reference gene candidates and a small selection of widely used reference genes, on 44 novel biological samples. Here, the fundamental issue is the need for a reference gene to perform the measurements, but which reference gene to use is the problem to be addressed. In Vandesompele et al. (2002) a method is presented, ranking the candidates based on the fact that ratios of abundance of reference genes in different samples should be constant. In the current paper, an extension is introduced, performing bootstrapping to assess the credibility of the ranks. Briefly, a novel reference gene, LRP10, is ranked best in 71% of the cases and the ratio of LRP10 and CLN3 show considerably less both systematic and non-systematic variation than the best ratios of widely used reference genes.

An interesting alternative to the approach of Vandesompele et al. (2002) has been presented by Andersen et al. (2004). Here a variance component model is used to assess the variability of the different reference gene candidates from realtime RT-PCR data.

Comments on Paper III:
Weighted analysis of paired microarray experiments.

In Paper III (Kristiansson et al., 2005), a novel model for the analysis of paired microarray experiments is introduced, titled *Weighted analysis of paired microarray experiments* (WAME). In previous papers (Lönnstedt and Speed, 2002; Smyth, 2004), parametric models explicitly modelling the prior distribution for the gene-specific variances by the inverse gamma distributions have been proposed. There the large number of genes is used to determine the hyperparameters for the prior according to the empirical Bayes principle. In the current paper, the model is extended to catch array-wise variations in quality, allowing for: (i) different variances for different repetitions (technical or biological) and (ii) correlations catching e.g. shared sources of variation. Methods are developed for estimating the covariance matrix and the hyperparameter for the prior of the gene-specific scaling components and a likelihood ratio test is derived for the differential expression of each gene. The results from some simulated and three real datasets are presented. On simulated data, WAME is performing at least as well as the existing methods, even when data is deviating from the model. In WAME, the structure of the data is utilised in a novel way, using the large number of genes that are non-differentially expressed to estimate the covariance matrix, which is complicated by the heavy tails of the distribution for the gene-specific variance scales.

References

C.L. Andersen, J.L. Jensen, and T.F. Ørntoft. Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. *Cancer Research*, 64(15):5245–5250, 2004.

P. Baldi and A.D. Long. A bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics*, 17(6):509–519, 2001.

M. Benson, L. Carlsson, M. Adner, M. Jernås, M. Rudemo, A. Sjögren, P.A. Svensson, R. Uddman, and Cardell L.O. Gene profiling reveals increased expression of uteroglobin and other anti-inflammatory genes in glucocorticoid-treated nasal polyps. *Journal of Allergy and Clinical Immunology*, 113(6):1011–1284, 2004.

B.M. Bolstad. *Low-level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization*. PhD thesis, University of California, Berkeley, California, 2004.

B. Efron, R. Tibshirani, J.D. Storey, and V. Tusher. Empirical Bayes Analysis of a Microarray Experiment. *Journal of the American Statistical Association*, 96(456):1151–1160, 2001.

B.G. Gabrielsson, L.E. Olofsson, A. Sjögren, M. Jernås, A. Elander, M. Lönn, M. Rudemo, and L.M.S. Carlsson. Evaluation of reference genes for studies of gene expression in human adipose tissue. *Obesity Research*, 13(4, to appear), 2005.

E. Kristiansson, A. Sjögren, M. Rudemo, and O. Nerman. Weighted analysis of paired microarray experiments. to be submitted, 2005.

I. Lönnstedt and T. Speed. Replicated microarray data. *Statistica Sinica*, 12 (1):31–46, 2002.

G. Smyth. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1), 2004.

V.G Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, 98(9):5116–5121, 2001.

J. Vandesompele, K.D. Preter, F. Pattyn, B. Poppe, N. van Roy, A.D. Paepe, and F. Speleman. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3(7), 2002.

Paper I

Gene profiling reveals increased expression of uteroglobin and other anti-inflammatory genes in glucocorticoid-treated nasal polyps

Mikael Benson, MD, PhD,[a,b] Lena Carlsson, MD, PhD,[c] Mikael Adner, PhD,[a] Margareta Jernås, BSc,[c] Mats Rudemo, PhD,[d] Anders Sjögren, BSc,[d] Per Arne Svensson, PhD,[c] Rolf Uddman, MD, PhD,[a] and Lars Olaf Cardell, MD, PhD[a] *Malmö and Gothenburg, Sweden*

Background: Treatment with local glucocorticoids (GCs) decreases symptoms and the size of nasal polyps. This might depend on the downregulation of proinflammatory genes, as well as the upregulation of anti-inflammatory genes.
Objective: We sought to identify GC-regulated anti-inflammatory genes in nasal polyps.
Methods: Affymetrix DNA microarrays were used to analyze the expression of 22,283 genes in 4 nasal polyps before and after local treatment with fluticasone (400 μg/d). Expression of uteroglobin and mammaglobin B was analyzed with real-time PCR in 6 nasal polyps and in nasal biopsy specimens from 6 healthy control subjects.
Results: Two hundred three genes had changed in expression in treated polyps, and 139 had known functions: 54 genes were downregulated, and 85 were upregulated. Genes associated with inflammation constituted the largest single functional group. These genes affected key steps in inflammation (eg, immunoglobulin production; antigen processing and presentation; and the chemoattraction and activation of granulocytes, T cells, and B cells). Several proinflammatory genes were downregulated. In contrast, some anti-inflammatory genes were upregulated. The gene that increased most in terms of expression was uteroglobin. This was confirmed with real-time PCR. By contrast, expression of uteroglobin was lower in untreated polyps than in healthy nasal mucosa. Immunohistochemical investigation showed staining of uteroglobin in the epithelium and in seromucous glands in control subjects and in nasal polyps.
Conclusion: Upregulation of anti-inflammatory genes, such as uteroglobin, might contribute to the effects of local treatment with GCs in nasal polyps. (J Allergy Clin Immunol 2004;113:1137-43.)

Key words: *Polyp, microarray, uteroglobin*

Abbreviations used
CCL: Chemokine (C-C motif) ligand
Ct: Cycle threshold
GC: Glucocorticoid
PDE: Phosphodiesterase

Nasal polyposis is often used as an in vivo model to study the effects of local treatment with glucocorticoids (GCs). Nasal polyps are characterized by an eosinophilic edematous stroma.[1] Immunohistochemical investigations have revealed mast cell degranulation and high local levels of IgE.[2] The effects of GCs in treating nasal polyps have not been fully elucidated. The downregulation of proinflammatory cytokines and adhesion molecules that attract and activate eosinophils has been demonstrated.[3] Recently, upregulation of the anti-inflammatory cytokine TGF-β was described in GC-treated nasal polyps.[4] This suggests that increased anti-inflammatory activity could contribute to the effects of GC treatment.

DNA microarrays consist of a matrix with attached DNA sequences that permit simultaneous analysis of the expression of thousands of genes. This provides unique opportunities to analyze the effects of GCs on a genome-wide scale. Studies of this kind have been performed on a variety of cell types.[5-9] They reveal that GCs not only downregulate proinflammatory mediators but also upregulate anti-inflammatory agents. However, gene expression profiles vary considerably in different cell types. This is consistent with previous data. For example, GCs induce apoptosis in eosinophils but inhibit apoptosis in neutrophils.[10] Different doses or types (ie, natural or synthetic) of GCs could also contribute to the variety of gene expression profiles. It is therefore of interest to examine the in vivo effects of pharmacologic doses of GCs on human tissue by using DNA microarrays. To our knowledge, no such studies exist, but microarray analyses of nasal mucosa from patients with allergic rhinitis with and without nasal polyps have been performed.[11,12]

In this study the effects of topical GC treatment on nasal polyps were examined with DNA microarrays measuring the expression of some 22,283 genes, with particular emphasis on anti-inflammatory genes.

From [a]the Laboratory for Clinical and Experimental Allergy Research, Department of Oto-Rhino-Laryngology, Malmö University Hospital; [b]the Pediatric Allergy Research Group, Queen Silvia Children's Hospital, Gothenburg; [c]the Research Centre for Endocrinology and Metabolism, Sahlgrenska University Hospital, Gothenburg; and [d]the Department of Mathematical Statistics, Chalmers University of Technology, Gothenburg.
Supported by the Swedish Medical Research Council, Swegene; the Swedish Heart and Lung Foundation; the Lundberg Foundation and the Swedish Asthma and Allergy Foundation; and the Knut and Alice Wallenberg Foundation.
Received for publication August 11, 2003; revised February 16, 2004; accepted for publication February 18, 2004.
Reprint requests: Mikael Benson, MD, PhD, Pediatric Allergy Research Group, Queen Silvia Children's Hospital, SE-416 85 Gothenburg, Sweden.
0091-6749/$30.00

doi:10.1016/j.jaci.2004.02.028

METHODS

Patients

Six male patients with bilateral nasal polyposis requiring surgical intervention (median age, 48 years; age range, 41-54 years) were included in this study (Table E1 in the Journal's Online Repository at www.mosby.com/jaci). In addition, nasal mucosal biopsy specimens were obtained from 6 healthy control subjects for examination with real-time PCR. Their median age was 31 years (range, 24-47 years), and 2 were women. Nasal polyposis was identified on the basis of clinical symptoms and the visualization of polyps by means of anterior rhinoscopy. A full ear, nose, and throat examination and a skin prick test were performed before inclusion. Patients with cystic fibrosis and ciliary dyskinesia were excluded from the study along with subjects with a history of concurrent purulent nasal infection in the 6 weeks before the study or any kind of nasal surgery during the last year. None of the patients or healthy control subjects was an active smoker or subjected to passive smoke exposure on a regular basis. None of the patients or healthy control subjects had asthma that required continuous medication. Skin prick tests were performed as previously described.[11] None of the healthy control subjects but 2 of the patients had a positive skin test response to birch pollen, grass pollen, or both, with a history of intermittent allergic rhinitis. These patients participated in the study during the autumn-winter (outside the pollen season).

Local treatment with GCs was withheld for a minimum of 6 weeks before the study. After this running-in period, a first set of polyps was surgically removed after topical application of local anesthesia containing lidocainhydrochloride-nafazoline (34 mg/mL + 0.17 mg) for about 20 minutes. No other surgical procedures were performed on this occasion. One week later, the patients were re-examined by the surgeon, and fluticasone, 200 μg twice daily, was initiated. After 6 weeks on this course, a new set of polyps was removed.

All patients were recruited through physician referrals. This study was approved by the ethics committee of the University of Lund, and informed written consent was obtained from all subjects.

DNA microarray analysis, quantitative real-time PCR, and immunohistochemistry

For a detailed description, see the e-text in the Journal's Online Repository. Briefly, nasal polyps from each patient before and after GC treatment were analyzed by using duplicate DNA microarrays measuring the expression of 22,283 genes (HuGe U133A GeneChip; Affymetrix, Santa Clara, Calif). Nasal mucosal biopsy specimens, macrophages, and adipose tissue were obtained and analyzed as previously described.[11,13,14]

Data analysis

Data analysis was performed in 2 steps. First, as an internal control, the expression levels of genes previously described as GC-regulated were examined. The selection of GC-regulated genes was based on a recent review.[15] Second, an open search was performed for genes that differed between nasal polyps before and after GC treatment. This search was based on a statistical method, as described below. The identities of the analyzed transcripts were verified by means of BLAST analysis (http://www.ncbi.nlm.nih.gov/BLAST/) and by means of database searches in the ExPASy Molecular Biology Server (http://www.expasy.ch).

Statistical analysis

The Spearman rank correlation test was used to analyze correlations between duplicate DNA microarrays.

Step 1. The Wilcoxon signed-rank test was used to compare microarray expression levels before and after GC treatment for 30 genes previously found to be downregulated by GCs.

Step 2. Expression values were transformed to cube roots to find out which genes were changed in expression levels,[16] and for each gene, a score value was computed as a modified t statistic: $z_g = \bar{y}_{gd}/(s_g + a_0)$. Here $\bar{y}_{gd}$ is the difference between averages of cube root values after and before GC treatment for n patients. Similarly, S_g is the SE corresponding to the n patients, and a_0 is a regularizing constant chosen as the 90th percentile for all S_g values.[17] To determine thresholds z_{min} and z_{max}, which give rise to small or large z_g values, a reference distribution was formed corresponding to a variable, z_{0g}. This variable was constructed analogously to the variable z_g but using the differences between the A and B arrays instead of differences between levels before and after GC treatment.[17] The differences can be created in 4 ways for each individual for each gene. The distribution of z_{0g} was defined as the average of the distribution for all the corresponding combinations. The values of the thresholds z_{min} and z_{max} were set according to that z_{0g} distribution. It should be noted that in addition to the effect of GC treatment, the distribution of z_g differs from the distribution of z_{0g} also by including effects of different polyps and of the time span of 6 weeks between polyp acquirement. The levels for z_{min} and z_{max} were therefore stringently chosen as the lower and upper 0.1% levels.

RESULTS

General characteristics of the gene expression data

Nasal polyps before and after treatment with GCs were analyzed with duplicate DNA microarrays that measured the expression of 22,283 genes. The quality of the microarray data was assessed during various control experiments. This led to the exclusion of one of the patients (see the Methods section in the Journal's Online Repository at www.mosby.com/jaci). The reproducibility of the gene expression data for the remaining 3 patients was supported by good agreement between the duplicate microarrays. The mean ± SEM correlation coefficients *(r)* for the duplicate expression levels were 0.95 ± 0.02 for the untreated polyps and 0.92 ± 0.07 for the treated polyps.

Data analysis

Step 1*: Analysis of genes previously described as downregulated by GCs.* A group of 30 genes earlier found to be downregulated by GCs was defined on the basis of a recent review to compare the effects of GCs in this study with previously described effects.[15] This served as an internal control. Twenty of the 30 genes differed more than 10% (comparing levels before and after GC treatment): 18 of these had a lower expression and 2 had a higher expression after treatment (P = .001, Wilcoxon signed-rank test; Fig 1).

Step 2*: Unbiased search for GC-regulated genes.* Two hundred three genes differed in expression in nasal polyps before and after GC treatment (Tables E2 and E3 in the Journal's Online Repository at www.mosby.com/jaci). The median ratios between expression levels before and after treatment were 0.46 for downregulated genes and 2.24 for upregulated genes. By using a public database (http://www.ncbi.nlm.nih.gov/LocusLink/), the genes were organized in functional

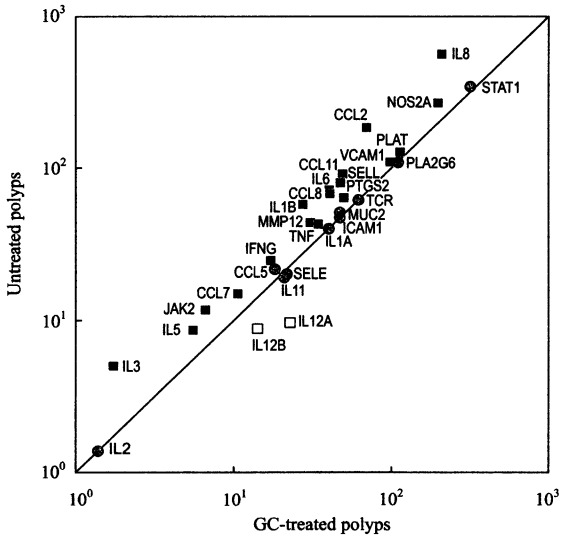


FIG 1. Expression levels of genes previously described as downregulated[15] by GCs in nasal polyps before and after treatment with a local GC. The *diagonal line* is the line of identity. Ten of the 30 genes differed less than 10% *(circles)*, and of the remaining 20, 2 had a higher *(open boxes)* and 18 had a lower *(filled boxes)* expression after treatment ($P = .001$, Wilcoxon signed-rank test). *CCL,* Chemokine (C-C motif) ligand; *ICAM1,* intercellular adhesion molecule 1 (CD54), human rhinovirus receptor; *IFNG,* IFN-γ; *JAK2,* Janus kinase 2 (a protein tyrosine kinase); *MCP,* monocyte chemotactic protein; *MMP,* matrix metalloproteinase 12; *MUC2,* mucin 2, intestinal, tracheal; *NOS2A,* nitric oxide synthase 2A, inducible, hepatocytes; *PLA2G6,* phospholipase A2, group VI, cytosolic, calcium independent; *PLAT,* plasminogen activator, tissue; *PTGS2,* prostaglandin-endoperoxide synthase 2; *SELE,* selectin E (endothelial adhesion molecule 1); *SELL,* selectin L (lymphocyte adhesion molecule 1); *STAT,* signal transducer and activator of transcription 1, 91 kd; *TCR,* T-cell receptor; *TNF,* tumor necrosis factor (TNF superfamily, member 2); *VCAM1,* vascular cell adhesion molecule 1.

TABLE I. Inflammatory genes with lower expression after treatment with GCs

Accession no.	Description	Ratio,* reference†
NM_004942.2	Defensin, β2	0.13
NM_006538.1	BCL2-like 11 (apoptosis facilitator)	0.20
M85276	Granulysin isoform NKG5	0.23
NM_002600.1	Phosphodiesterase 4B	0.25 ↓[18]
NM_006128.1	Bone morphogenetic protein 1	0.25
NM_000584.1	IL-8	0.37 ↓[19]↓[15]↔[20]
NM_004049.1	BCL2-related protein A1	0.37
L10343	Elafin	0.38
NM_004073.1	Cytokine-inducible kinase 17	0.40
U88321.1	Chemokine ligand 19	0.42
NM_000954.1	Prostaglandin D2 synthase	0.45
U64094.1	Soluble type II IL-1 receptor	0.46 ↑[38]↓[39]
NM_001803.1	CD52	0.46
NM_005564.1	Lipocalin 2	0.48
AF003934.1	Prostate differentiation factor	0.48
Y13710	Activated macrophage-specific CC chemokine 1	0.55
BF342851	Peroxidasin	0.58
AI743792	Sialyltransferase 1	0.62
NM_001908.1	Cathepsin B	0.69
NM_002121.1	Major histocompatibility complex, class II, DP β1	0.70
M63438.1	Immunoglobulin-rearranged γ chain	0.77
AA573862	Major histocompatibility complex, class I, A	0.79 ↓[21]

*Ratio between gene expression levels in nasal polyps before and after treatment with GCs (with genes ordered by ratios). Each microarray experiment was performed in duplicate (n = 4 for both untreated and treated polyps).
†References to previous studies. *Arrows* indicate increased, decreased, or unchanged expression of each gene.

groups: 139 of the 203 genes had known functions, with 54 being downregulated and 85 being upregulated. Genes associated with inflammation constituted the largest single functional group, comprising 32% of the total. Genes involved in transcription and translation (12%) and metabolism (12%) were the other major functional categories. Downregulated inflammatory genes are listed in Table I,[18-21] and upregulated inflammatory genes are listed in Table II.

Several proinflammatory genes were downregulated, such as genes regulating the influx of leukocytes (prostaglandin D_2 synthase, chemokine ligand 19, activated macrophage-specific CC chemokine 1, and IL-8), antigen processing (eg, cathepsin B), antigen presentation (MHC class I and II), phagocytosis (peroxidasin), and the general activation of inflammatory cells (phosphodiesterase 4B) and specific cells, such as T lymphocytes (granulysin), B lymphocytes (sialyltransferase 1 and immunoglobulins), macrophages (β-site APP-cleaving enzyme 2), and neutrophils (defensin β2 and lipocalin 2). Moreover, CD52, a surface marker on eosinophils, T cells, and B cells, was downregulated.

In contrast, some anti-inflammatory genes were upregulated. Uteroglobin was the gene the expression of which increased most after treatment (ie, 4.2-fold compared with before treatment). Uteroglobin is one of the most abundant proteins in nasal secretions and has wide-ranging anti-inflammatory properties. To independently confirm the validity of the microarray data on relative mRNA expression levels, we used TaqMan real-time quantitative RT-PCR. This was performed on nasal polyps from 6 patients before and after treatment. The levels of uteroglobin were related to the expression of the reference gene β-actin, and the Δ cycle threshold (Ct) values were 1.46 ± 0.61 and -1.26 ± 0.82 before and after treatment, respectively. After ΔΔCt analysis, the data demonstrated a 5.56 ± 2.58−fold increase in the expression of the uteroglobin gene in 6 patients treated with GCs ($P = .0071$). Mammaglobin B, another member of the uteroglobin family, was also among the upregulated genes, 2.0-fold compared with before treatment. This was confirmed by means of real-time PCR analysis of nasal polyps from 6 patients before and after treatment (ΔCt

TABLE II. Inflammatory genes with higher expression after treatment with GCs

Accession. no.	Description	Ratio,* reference†
NM_003357.1	Uteroglobin	4.2 ↑[34]
NM_004038.1	Amylase, α 1A	3.3
M64497.1	Nuclear receptor subfamily 2, group F	3.3
M15872.1	Glutathione S-transferase A2	3.3
X69397.1	CD24	3.2
NM_013230.1	CD24	3.2
S73751.1	Monocyte macrophage serine esterase 1	3.1
NM_000635.1	HLA regulatory factor X, 2	2.9
J05064.1	Complement component C6	2.6
AF054817.1	CD84	2.5
NM_005060.1	RAR-related orphan receptor C	2.5
NM_006992.1	B7 protein	2.3[40]
NM_002231.2	CD82	2.2
NM_003012.2	Secreted apoptosis-related protein	2.0
BF669455	CD164	1.8
NM_004079.1	Cathepsin S	1.7
NM_001140.1	Arachidonate 15-lipoxygenase	1.7
NM_021777.1	A disintegrin and metalloproteinase domain 28	1.7
BF793951	Serine protease 15	1.6
AF009616.1	CASP8 and FADD-like apoptosis regulator	1.5
AF182645.1	IK-cytokine, downregulator of HLA	1.5

*Ratio between gene expression levels in nasal polyps before and after treatment with GCs (with genes ordered by ratios). Each microarray experiment was performed in duplicate (n = 4 for both untreated and treated polyps).
†Reference to a previous study. The *arrow* indicates increased expression of the gene.

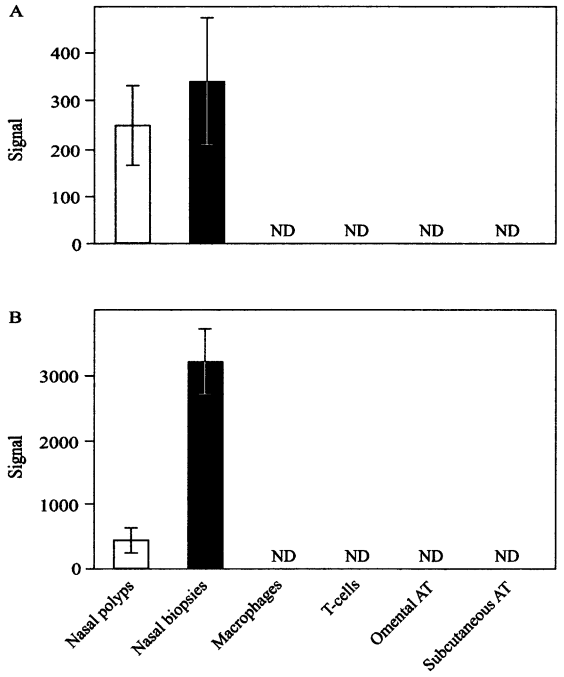


FIG 2. Mean ± SEM expression levels of mammaglobin B **(A)** and uteroglobin **(B)** in nasal polyps *(open bars)* and nasal biopsy specimens *(filled bars)* from healthy control subjects and in various control human cells and tissues. The expression levels in nasal polyps are derived from experiments with HuGeU133A Gene Chips and the other expression levels from experiments with HuGe95A GeneChips. *ND,* Not detectable; *AT,* adipose tissue; *SC,* subcutaneous.

values were 4.54 ± 0.93 and 2.26 ± 1.54, respectively), which rendered a 3.86 ± 2.16−fold increase after GC treatment (P = .0108).

In contrast to uteroglobin, the function of mammaglobin B is not known. Examination of DNA microarray analyses of various cells and tissues performed at our laboratory showed that both mammaglobin B and uteroglobin were only expressed in nasal polyps and mucosa (Fig 2). Interestingly, the expression of uteroglobin was higher in normal nasal mucosa than in untreated nasal polyps. This suggests that decreased uteroglobin could contribute to the disease process in nasal polyposis. However, the nasal polyp values were obtained with a more recent version of the DNA microarray GeneChip and not directly comparable. Therefore real-time PCR analysis of uteroglobin was performed in nasal biopsy specimens from 6 healthy control subjects. The expression of uteroglobin showed a 0.09 ± 0.03−fold decrease in untreated nasal polyps compared with healthy nasal mucosa (ΔCt values were 1.46 ± 0.61 and −2.02 ± 0.30, respectively; ie, 11 times lower in the polyps, P = .0005). The corresponding fold change for mammaglobin B was 1.05 ± 1.83 (ΔCt values were 4.54 ± 0.89 and 4.61 ± 0.82, respectively; P = .953).

Immunohistochemical investigation was performed to examine the expression of uteroglobin in nasal biopsy specimens from healthy control subjects and in untreated nasal polyps. In the nasal mucosa from healthy individuals, an intense immunostaining of uteroglobin was detected in the acini of seromucous glands and in epithelial cells. In specimens from nasal polyps, a similar distribution was seen, although the staining of epithelial cells was less intense and a smaller number of cells displayed immunoreactivity (Fig 3).

Other upregulated genes included arachidonate 15-lipoxygenase, an enzyme that induces production of lipoxins, a class of anti-inflammatory eicosanoids; IK-cytokine, which reduces MHC class II expression; and secreted apoptosis-related protein, which is proapoptotic. Two CD markers with anti-inflammatory properties were upregulated: CD24, which induces B-cell apoptosis, and CD164, which is cytoprotective and antiadhesive. Detoxifying agents that protect against inflammatory damage in

the airways were also upregulated (ie, glutathione S-transferase A2 and monocyte macrophage serine esterase 1).

DISCUSSION

In this study the expression of 22,283 genes was analyzed in nasal polyps before and after GC treatment. Of these genes, 203 differed in expression, and 139 had known functions. The largest functional group comprised genes related to inflammation. Several proinflammatory genes were downregulated. These affected key steps in inflammation, ranging from the influx of leukocytes to immunoglobulin production. Although most of these genes have not been previously described as GC regulated, their relevance is suggested by 2 of them being targets for specific anti-inflammatory therapy: phosphodiesterase (PDE) 4B and the CD52 antigen. PDEs inactivate cyclic AMP, and PDE4 is the predominant isoenzyme in inflammatory cells. Recently, a selective PDE4 inhibitor has been developed to treat chronic obstructive pulmonary disease.[22] CD52 is expressed on lymphocytes. Antibodies directed against CD52 are used as immunosuppressants in transplantation and autoimmune disease.[23] This suggests that identification of genes that are downregulated by GCs might help to find new therapeutic candidates. However, the effects of GCs might depend not only on downregulation of proinflammatory genes but also on upregulation of anti-inflammatory genes. This is the first study to confirm this in vivo on a genome-wide scale.

There are several methodological concerns in microarray studies. A large number of genes are analyzed in a small number of patients. This involves the risk of spurious findings. To reduce this risk, a statistical method that takes into account the variance of the data was used to identify differentially expressed genes, rather than a fixed cutoff point, such as a 2-fold change in expression.[16,17,24] The validity of the data in this study is supported by several control experiments that are part of the analytic protocol. Decreased expression of proinflammatory genes in this study could be caused by decreased influx of cells rather than a direct effect on gene expression. As an internal control, the effects of GCs in this study were compared with those described in a recent review.[15] The majority of the genes that were reported to be downregulated by GCs also had lower expression in this study. The finding that these genes included several related to eosinophil adhesion, chemotaxis, and activation (*VCAM1*, *CCL5*, *CCL11*, *IL3*, *IL5*, *IL8*, and *TNF*) was of particular interest. This agrees with previous in vivo studies indicating that GCs inhibit T_H2 cytokines in nasal polyps and allergic rhinitis[2,3,25-27] but not with some in vitro studies.[15,28] It is worth noting that not all genes can be characterized as either proinflammatory or anti-inflammatory. For example, T_H1 cytokines inhibit IgE-mediated allergic reactions but enhance type IV hypersensitivity. In this study the T_H1 cytokine *IFNG* did not change, and the T_H1-promoting cytokines *IL12A* and *IL12B* even tended to increase. This is in agreement with previous studies of nasal polyps and allergic disease showing either no change or an increase in *IFNG* after GC treatment.[4,8,25,26,29,30]

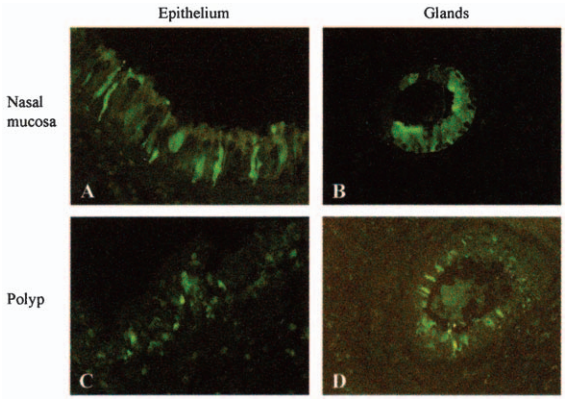


FIG 3. Expression of uteroglobin protein determined by means of immunohistochemistry in the nasal mucosa from one representative healthy control subject **(A** and **B)** and from one nasal polyp **(C** and **D)** at 200× magnification. Immunoreactivity for uteroglobin is found in the epithelium (Fig 3, *A*) and in the acini of seromucous glands (Fig 3, *B*). In nasal polyps uteroglobin immunofluorescence is found in the epithelium (Fig 3, *C*) and in glandular acini (Fig 3, *D*).

Little is known about the effects of GCs on anti-inflammatory genes in human tissue. Recently, increased expression of the anti-inflammatory gene *TGFB* was described in GC-treated nasal polyps.[4] This suggests that such genes could contribute to the beneficial effects of GCs. In this study of nasal polyps, uteroglobin was the gene that increased most after GC treatment. This was confirmed with real-time PCR analysis. Uteroglobin, or Clara cell 10-kd protein, is known to be secreted by Clara cells in the lungs. To our knowledge, uteroglobin has not been previously described in the nasal mucosa. It has wide-ranging anti-inflammatory effects (ie, the inhibition of leukocyte chemotaxis, phospholipase A2, and proinflammatory cytokines, as well as protease activity). The relevance of these anti-inflammatory effects is supported by decreased local expression of uteroglobin protein in asthma.[31] Peptides derived from uteroglobin are among the most potent anti-inflammatory agents identified to date.[32] Recently, a peptide of this kind was successfully tested in an animal model of allergic conjunctivitis.[33] Experimental data show that GCs induce the expression of uteroglobin.[34] To our knowledge, this is the first report to confirm this in vivo. Because treatment with GCs is known to decrease symptoms and the size of nasal polyps, as well as eosinophil infiltration, it is possible that uteroglobin might contribute to these effects. Interestingly, comparisons with previous microarray studies indicated lower expression of uteroglobin in untreated nasal polyps compared with in nasal mucosa from healthy control subjects. This was confirmed with real-time PCR that showed that the expression levels were 11 times lower in the polyps.

Immunohistochemical investigation of nasal biopsy specimens from healthy control subjects revealed intense immunostaining of uteroglobin in the acini of seromucous

glands and in epithelial cells. A similar distribution was seen in untreated nasal polyps, but the staining of epithelial cells was less intense, and a smaller number of cells displayed immunoreactivity. Taken together, these novel findings indicate that uteroglobin might have a role in regulating inflammation in the nasal mucosa. Further studies are needed to examine whether altered expression of uteroglobin contributes to the pathogenesis of nasal polyposis, as well as the beneficial effects of GCs in this disease.

Mammaglobin B, another member of the uteroglobin family, also increased in nasal polyps after treatment. In a recent microarray study of nasal mucosa from patients with allergic rhinitis with and without nasal polyps, increased expression of mammaglobin B was found in the patients with polyps.[12] Interestingly, in our study expression of mammaglobin B did not significantly differ when healthy nasal mucosa was compared with untreated nasal polyps. This suggests that further studies are needed to examine the role of mammaglobin B in nasal polyposis and allergic rhinitis. The function of mammaglobin has not been defined. Different observations suggest that mammaglobin B and uteroglobin might be functionally related. The 2 genes display sequence homology, and an examination of DNA microarray analyses of various cells and tissues performed at our laboratory revealed that the genes were only expressed in nasal polyps and mucosa. In addition, the expression of both genes increased after GC treatment.

The upregulated genes included another gene with wide-ranging anti-inflammatory properties, arachidonate 15-lipoxygenase. This is an enzyme that induces production of lipoxins, a class of eicosanoids that are important stop signals for inflammatory reactions.[35] Other upregulated anti-inflammatory genes included IK-cytokine, which decreases MHC class II expression. This is consistent with the lower expression of MHC class II that was demonstrated in the GC-treated polyps. Similarly, the increased expression of CD24, which induces B-cell apoptosis,[36] agrees with the reduced expression of genes expressed by B cells, such as CD52, sialyltransferase 1, and immunoglobulins.

The enhancement of anti-inflammatory genes could therefore contribute to the beneficial effects of GCs in vivo. Not all genes matched the simple concept of GC downregulating proinflammatory genes and upregulating anti-inflammatory genes. For example, the soluble type 2 IL-1 receptor decreased after GC treatment. Soluble cytokine receptors might act as both enhancers or inhibitors, depending on the relative concentration between the soluble receptor and its ligand.[37] Previous experimental data indicate that GCs might both increase and reduce the expression of the soluble type 2 IL-1 receptor.[38,39] Increased expression of B7 is another example of the complex effects of GC. B7 is expressed on antigen-presenting cells and might either activate or inhibit T cells, depending on which other costimulatory molecules are expressed.[40]

To summarize, GCs suppress the expression of many proinflammatory genes and enhance some anti-inflammatory genes. The characterization of the balance between these genes might contribute to an understanding of the effects of GCs. Profiling gene expression after GC treatment might also help to identify therapeutic candidates.

REFERENCES

1. Mygind N, Dahl R, Bachert C. Nasal polyposis, eosinophil dominated inflammation and allergy. Thorax 2000;55(suppl 2):S79-83.
2. Bachert C, Gevaert P, Holtappels G, Johansson SG, van Cauwenberge P. Total and specific IgE in nasal polyps is related to local eosinophilic inflammation. J Allergy Clin Immunol 2001;107:607-14.
3. Hamilos DL, Thawley SE, Kramper MA, Kamil A, Hamid QA. Effect of intranasal fluticasone on cellular infiltration, endothelial adhesion molecule expression, and proinflammatory cytokine mRNA in nasal polyp disease. J Allergy Clin Immunol 1999;103:79-87.
4. Mastruzzo C, Greco LR, Nakano K, Nakano A, Palermo F, Pistorio MP, et al. Impact of intranasal budesonide on immune inflammatory responses and epithelial remodeling in chronic upper airway inflammation. J Allergy Clin Immunol 2003;112:37-44.
5. Webster JC, Huber RM, Hanson RL, Collier PM, Haws TF, Mills JK, et al. Dexamethasone and tumor necrosis factor-alpha act together to induce the cellular inhibitor of apoptosis-2 gene and prevent apoptosis in a variety of cell types. Endocrinology 2002;143:3866-74.
6. Yoshida NL, Miyashita T, Mami U, Yamada M, Reed JC, Sugita Y, et al. Analysis of gene expression patterns during glucocorticoid-induced apoptosis using oligonucleotide arrays. Biochem Biophys Res Commun 2002;293:1254-61.
7. Maurer M, Trajanoski Z, Frey G, Hiroi N, Galon J, Willenberg HS, et al. Differential gene expression profile of glucocorticoids, testosterone, and dehydroepiandrosterone in human cells. Horm Metab Res 2001;33:691-5.
8. Obexer P, Certa U, Kofler R, Helmberg A. Expression profiling of glucocorticoid-treated T-ALL cell lines: rapid repression of multiple genes involved in RNA-, protein- and nucleotide synthesis. Oncogene 2001;19:4324-36.
9. Tonko M, Ausserlechner MJ, Bernhard D, Helmberg A, Kofler R. Gene expression profiles of proliferating vs. G1/G0 arrested human leukaemia cells suggest a mechanism for glucocorticoid-induced apoptosis. FASEB J 2001;15:693-9.
10. Meagher LC, Cousin JM, Seckl JR, Haslett C. Opposing effects of glucocorticoids on the rate of apoptosis in neutrophilic and eosinophilic granulocytes. J Immunol 1996;156:4422-8.
11. Benson M, Cardell LO, Carlsson B, Jernås M, Reinholdt J, Svensson PA, et al. DNA microarrays to study gene expression in allergic airways. Clin Exp Allergy 2002;30:1-8.
12. Fritz SB, Terrell JE, Conner ER, Kukowska-Latallo JF, Baker JR. Nasal mucosal gene expression in patients with allergic rhinitis with and without nasal polyps. J Allergy Clin Immunol 2003;112:1057-63.
13. Gabrielsson BG, Johansson JM, Jennische E, Jernas M, Itoh Y, Peltonen M, et al. Depot-specific expression of fibroblast growth factors in human adipose tissue. Obes Res 2002;10:608-16.
14. Ohlsson BG, Englund MC, Karlsson AL, Knutsen E, Erixon C, Skribeck H, et al. Oxidized low density lipoprotein inhibits lipopolysaccharide-induced binding of nuclear factor-kappaB to DNA and the subsequent expression of tumor necrosis factor-alpha and interleukin-1beta in macrophages. J Clin Invest 1996;98:78-89.
15. Webster JI, Tonelli L, Sternberg EM. Neuroendocrine regulation of immunity. Annu Rev Immunol 2002;20:125-6.
16. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Nat Acad Sci 2001;98:5116-21.
17. Efron B, Tibshirani R, Storey JD, Tusher V. Empirical Bayes analysis of a microarray experiment. J Am Stat Assoc 2001;96:1151-60.
18. Delgado M, Fernández-Alfonso SM, Fuentes J. Effect of adrenaline and glucocorticoids on monocyte cAMP-specific phosphodiesterase (PDE4) in a monocytic cell line. Arch Dermatol Res 2002;4:190-7.

19. Sim TC, Reece LM, Hilsmeier KA, Grant JA, Alam R. Secretion of chemokines and other cytokines in allergen-induced nasal responses: inhibition by topical steroid treatment. Am J Respir Crit Care Med 1995; 152:927-33.
20. Benson M, Strannegård IL, Wennergren G, Strannegård Ö. Interleukin-5 and interleukin-8 in relation to eosinophils and neutrophils in nasal fluids from school children with seasonal allergic rhinitis. Pediatr Allergy Immunol 1999;10:178-85.
21. Cancedda MHC, Filaci G, Puppo F, Ghio M, Contini P, Indiveri F. Immune homeostasis requires several biologic factors including glucocorticoid hormones. Ann N Y Acad Sci 2002;966:49-63.
22. Gamble E, Grootendorst DC, Brightling CE, Troy S, Qui Y, Zhu J, et al. Anti-inflammatory effects of the phosphodiesterase 4 inhibitor cilomilast (Ariflo) in COPD. Am J Respir Crit Care Med 2003. [Epub ahead of print].
23. Hale G. Alemtuzumab in stem cell transplantation. Med Oncol 2002; 19(suppl):S33-47.
24. Terry Speed, editor. Statistical analysis of gene expression microarray data. New York: Chapman & Hall/CRC; 2003.
25. Benson M, Strannegård IL, Wennergren G, Strannegård Ö. Cytokines in nasal fluids from school children with seasonal allergic rhinitis. Pediatr Allergy Immunol 1997;8:143-9.
26. Benson M, Strannegard I-L, Strannegard O, Wennergren G. Topical steroid treatment of allergic rhinitis decreases nasal fluid TH2 cytokines, eosinophils, eosinophil cationic protein, and IgE but has no significant effect on IFN-gamma, IL-1beta, TNF-alpha, or neutrophils. J Allergy Clin Immunol 2000;106:307-12.
27. Schwiebert LM, Beck LA, Stellato C, Bickel CA, Bochner BS, Schleimer RP. Glucocorticosteroid inhibition of cytokine production: relevance to antiallergic actions. J Allergy Clin Immunol 1996;97:143-52.
28. Ramirez F, Fowell DJ, Puklavec M, Simmonds S, Mason D. Glucocorticoids promote a Th2 cytokine response by CD4+ T cells in vitro. J Immunol 1996;156:2406-12.
29. Robinson D, Hamid Q, Ying S, Bentley A, Assoufi B, Durham S, et al. Prednisolone treatment in asthma is associated with modulation of bronchoalveolar lavage cell interleukin-4, interleukin-5, and interferon-gamma cytokine gene expression. Am Rev Respir Dis 1993;148:401-6.
30. Henriksson G, Norlander T, Forsgren J, Stierna P. Effects of topical budesonide treatment on glucocorticoid receptor mRNA down-regulation and cytokine patterns in nasal polyps. Am J Rhinol 2001;15:1-8.
31. Singh G, Katyal SL. Clara cell proteins. Ann N Y Acad Sci 2000;923: 43-58.
32. Miele L. New weapons against inflammation: dual inhibitors of phospholipase A2 and transglutaminase. J Clin Invest 2003;111:19-21.
33. Sohn J, Kim TI, Yoon YH, Kim JY, Kim SY. Novel transglutaminase inhibitors reverse the inflammation of allergic conjunctivitis. J Clin Invest 2003;111:121-8.
34. Berg T, Cassel TN, Schwarze PE, Nord M. Glucocorticoids regulate the CCSP and CYP2B1 promoters via C/EBPbeta and delta in lung cells. Biochem Biophys Res Commun 2002;293:907-12.
35. Levy D, Clish CB, Schmidt B, Gronert K, Serhan CN. Lipid mediator class switching during acute inflammation: signals in resolution. Nat Immunol 2001;2:612-9.
36. Suzuki T, Kiyokawa N, Taguchi T, Sekino T, Katagiri YU, Fujimoto J. CD24 induces apoptosis in human B cells via the glycolipid-enriched membranedomains/rafts-mediated signaling system. J Immunol 2001; 166:5567-77.
37. Heaney ML, Golde DW. Soluble receptors in human disease. J Leukoc Biol 1998;64:135-46.
38. Re F, Muzio RM, De N, Polentarutti JG, Giri A, Mantovani A, et al. The type II "receptor" as a decoy target for interleukin 1 in polymorphonuclear leukocytes: characterization of induction by dexamethasone and ligand binding properties of the released decoy receptor. J Exp Med 1994;179:739-44.
39. Daun JM, Ball RW, Cannon JG. Glucocorticoid sensitivity of interleukin-1 agonist and antagonist secretion: the effects of age and gender. Am J Physiol Integr Comp Physiol 2000;278:R855-62.
40. Lohr J, Knoechel B, Jiang S, Sharpe AH, Abbas AK. The inhibitory function of B7 costimulators in T cell responses to foreign and self-antigens. Nat Immunol 2003;4:664-9.

Paper II

rich3/z78-oby/z78-oby/z7800405/z780042d05a | thistled | S=4 | 2/24/05 | 10:37 | Art: 04-063 | Input-dlw

Short Communication

Evaluation of Reference Genes for Studies of Gene Expression in Human Adipose Tissue

Britt G. Gabrielsson, Louise E. Olofsson,* Anders Sjögren,‡ Margareta Jernås,* Anna Elander,† Malin Lönn,* Mats Rudemo,‡ and Lena M. S. Carlsson**

Abstract

GABRIELSSON, BRITT G., LOUISE E. OLOFSSON, ANDERS SJÖGREN, MARGARETA JERNÅS, ANNA ELANDER, MALIN LÖNN, MATS RUDEMO, AND LENA M. S. CARLSSON. Evaluation of reference genes for studies of gene expression in human adipose tissue. *Obes Res.* 2005;13:?–?.

Objective: The aim of this study was to evaluate reference genes for expression studies of human adipose tissue.

Research Methods and Procedures: Using 52 human adipose tissue expression profiles (HU95), 10 putative reference genes with the lowest variation in expression levels were selected for further studies. Expression stability of these 10 novel and 5 previously established reference genes was evaluated by real-time reverse transcriptase-polymerase chain reaction analysis. For this purpose, 44 adipose tissue biopsies from 27 subjects were chosen to include a wide range of parameters such as sex, age, BMI, depot origin, biopsy procedure, and effects of nutrition.

Results: *LRP10* was identified as the gene with the least variation in expression levels. The frequently used reference genes *RPLP0, 18S rRNA, PPIA, ACTB,* and *GAPD* were ranked as 4, 6, 7, 8, and 10, respectively.

Discussion: Our results suggest that *LRP10* is a better choice as reference for expression studies of human adipose tissue compared with the most frequently used reference genes.

Key words: housekeeping, bootstrapping, LRP10, β-actin, GAPD

Received for review February 20, 2004.

Accepted in final form January 14, 2005.

The costs of publication of this article were defrayed, in part, by the payment of page charges. This article must, therefore, be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

*Research Center for Endocrinology and Metabolism, Department of Internal Medicine and †Department of Plastic Surgery, the Sahlgrenska Academy, Göteborg, Sweden and ‡Department of Mathematical Statistics, Chalmers University of Technology, Göteborg, Sweden.

Address correspondence to Britt Gabrielsson, Research Center for Endocrinology and Metabolism, Gröna Stråket 8, Sahlgrenska University Hospital, S-413 45 Göteborg, Sweden. E-mail: britt.gabrielsson@medic.gu.se

Introduction

A survey of 40 studies published since 2001 shows that, in 70% of the papers, *ACTB*, *GAPD*, or *18S rRNA* were used as reference genes for reverse transcriptase-polymerase chain reaction (RT-PCR)[1] measurements of gene expression in human adipose tissue or adipocytes. However, the expression of these genes has been reported to vary considerably in other tissues and cells (1–3). In addition, we have previously observed that *ACTB* was regulated during diet-induced weight loss (4). This study was therefore performed to identify and evaluate novel reference genes for analysis of gene expression in human adipose tissue and to compare these with frequently used reference genes.

Research Methods and Procedures

Subjects and Samples

This study was approved by the Medical Ethics Committee at Göteborg University, and all participants gave written informed consent. All biopsies were taken after an overnight fast, immediately frozen in liquid nitrogen, and stored at −80 °C until analysis. Seventy-two biopsies from 27 subjects (6 men and 21 women) were used for the microarray expression profiling, and 44 biopsies from 27 subjects (14 men and 13 women) were used for the real-time RT-PCR analysis (supplemental data available online at http://www.obesityresearch.org). Procedures for RNA isolation and hybridization to the microarrays have been described previously (5,6).

Selection of Stably Expressed Genes in Human Adipose Tissue using Microarray Data

In total, 52 expression profiles of human adipose tissue from the above mentioned 72 biopsies were used for the selection (supplemental data available online at http://www.obesityresearch.org). To perform a preselection using the largest group of data ($n = 36$), the 50 genes with the lowest

[1] Nonstandard abbreviations: RT-PCR, reverse transcriptase-polymerase chain reaction; CV, coefficient of variation.

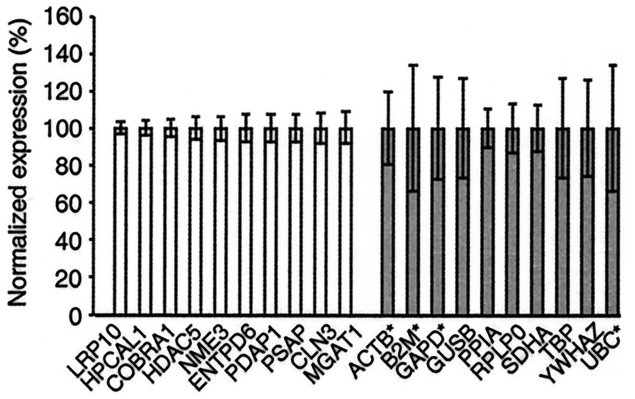


Figure 1: Expression stability of 10 novel (white bars) and 10 established (gray bars) reference genes in human adipose tissue analyzed by microarrays. The data are shown as normalized expression for each gene, and error bars indicate the CV, which was used to evaluate the expression stability of the genes (see Research Methods and Procedures).

coefficients of variation (CVs, %SD) were selected from expression profiles of paired subcutaneous and omental adipose tissue (5,6). Subsequently, all microarrays were used to create eight different groups representing different physiological variables of interest (supplemental data available online at http://www.obesityresearch.org). The average expression for each group was calculated, followed by computation of the CVs for the 50 genes across the eight groups. The 10 genes with the lowest CV were selected for subsequent studies (Figure 1; supplemental data available online at http://www.obesityresearch.org). F1

ACTB, *GAPD,* and *18S* were identified by PubMed searches as common reference genes for expression studies of human adipose tissue. Two other frequently used reference genes, *PPIA* and *RPLP0*, with low CV in the microarray data, were also included (Figure 1). The 5 established and 10 novel putative reference genes were evaluated by real-time RT-PCR analysis on individual samples.

Real-time RT-PCR

Eleven of the 15 genes were analyzed using predesigned TaqMan Assays-on-Demand (Applied Biosystems, Foster City, CA). Probe-primer sets for the four remaining genes, *COBRA1, ENTPD6, PDAP1,* and *HDAC5,* were designed using the Primer Express software v2.0 (Applied Biosystems). TaqMan Reverse Transcriptase reagents, TaqMan Universal PCR Master mix (Applied Biosystems), and reaction conditions were used according to the manufacturer's instructions (supplemental data available online at http://www.obesityresearch.org).

Ranking the Putative Reference Genes

Ranking of the selected reference genes was performed essentially as described by Vandesompele et al. (7). Briefly, for each gene, the gene expression ratio versus all other genes was calculated in each sample. Subsequently, for each pair of genes, a pairwise variation was defined as the SD of the pairwise log ratios for all samples, and for each gene, a gene instability measure was defined as the mean overall of the pairwise variations for that gene.

An iterative process was employed in the ranking procedure where genes were excluded stepwise. In each step, the gene with the highest gene instability measure was excluded, after which new gene instability measures were calculated using only the remaining genes. This procedure was repeated until only three genes remained. These genes were ranked as first, second, and third according to their gene instability measures.

In addition, we performed a bootstrap step to evaluate the certainty of the ranking. The ranking method was bootstrapped (8) by resampling with replacement from the original set of 44 sample files. The resampling procedure was repeated 10,000 times (supplemental data available online at http://www.obesityresearch.org). To check the robustness of the ranking procedure with respect to outliers, we also repeated the ranking with trimmed SDs, excluding the most outlying 10%, 20%, and 40% of log ratios in the computation of the SDs of the pairwise log ratios.

Results

Selection of Putative Reference Genes in Human Adipose Tissue from Microarray Data

Based on the analysis of data from 52 microarrays, the 10 genes with the least variability in expression levels were selected as possible novel reference genes. Figure 1 shows the normalized signal and the CVs for each of the 10 genes together with established reference genes represented on the same microarrays.

Evaluation of Gene Expression Stability

To evaluate the expression stability of the selected genes, we collected samples from both sexes with wide variation with respect to age (20 to 64 years), BMI (20.5 to 51.2 kg/m^2), nutrition (before diet, 8 weeks of diet, and 2 weeks refeeding after completed 16-week diet), depot (omental, subcutaneous), biopsy procedure (surgical, needle), and anesthesia (local, general). In total, 44 adipose tissue biopsies from 27 subjects were used for gene expression analysis by real-time RT-PCR. The assays for *PDAP1* and *MGAT1* yielded no or very low signals; consequently, only 13 genes were analyzed further.

The results from the ranking of the genes and the evaluation of the certainty of ranking by bootstrap procedure are shown in Figure 2. The figure shows that *LRP10* was ranked first in 71% of 10,000 bootstrap samples, followed by *CLN3* ranked first in 24% and second in 51% of the bootstrap samples. These results obtained by the bootstrap procedure F2

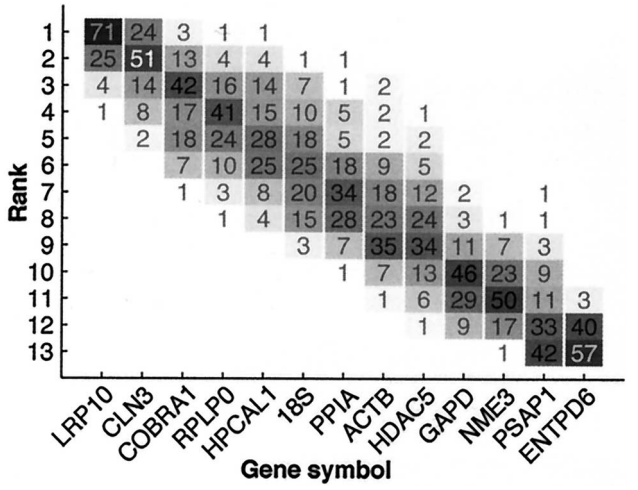


Figure 2: Empirical distribution of ranks shown in percentage for the 13 remaining genes analyzed by real-time RT-PCR in bootstrap of size 10,000. Stable genes have low ranks. The varying shading of numbers is only for improved readability of the figure.

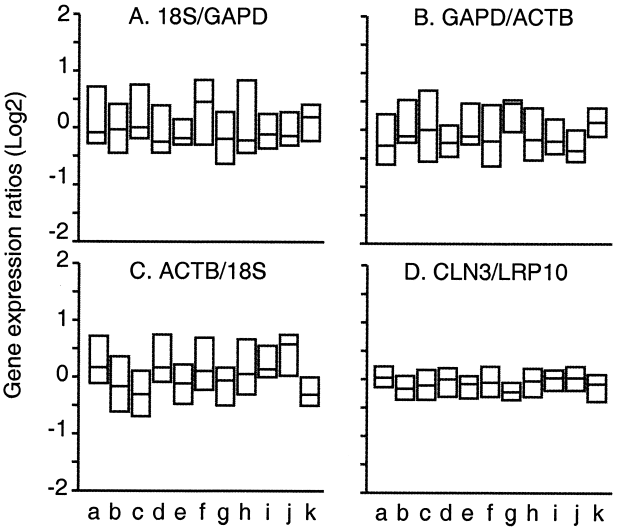


Figure 3: Variation in expression ratios of different gene combinations in human adipose tissue analyzed by real-time RT-PCR. (A) *18S/ACTB*. (B) *GAPD/ACTB*. (C) *ACTB/18S*. (D) *CLN3/LRP10*. The panels show box plots of the gene expression ratios in adipose tissue from (a) women (n = 12), (b) men (n = 13), (c) subjects with BMI <30 kg/m^2 (n = 12), (d) subjects with BMI ≥35 kg/m^2 (n = 13), (e) age <40 years (n = 13), (f) age ≥40 years (n = 12), (g) omental depot (n = 9), (h) subcutaneous depot (n = 7), (i) obese subjects, before weight loss (n = 8), (j) obese subjects, during weight loss (n = 8) and (k) obese subjects, after weight-loss (n = 4). The box plots show the quartiles of gene expression ratios where the expression of each gene was normalized before calculation of the ratios.

were in agreement with the original ranking. Furthermore, the ranking was also robust in that it was essentially unaffected by trimming away 10%, 20%, or 40% of the most outlying log ratios. The results suggest that *LRP10* is the most stably expressed gene.

Effects of Sex, BMI, Age, Depot, and Biopsy Procedure

To evaluate how the choice of reference gene affects the final results, the analyzed samples were grouped according to sex, BMI, age, depot origin, and effect of diet. Given that *18S*, *GAPD,* and *ACTB* are frequently used as reference genes, we chose to show the variation that was introduced when *18S* expression was related to *GAPD*, *GAPD* expression was related to *ACTB,* and *ACTB* expression was related to *18S* (Figure 3, A–C). The expression ratios of the two top ranked genes, *CLN3* and *LRP10*, are shown for comparison (Figure 3D). Figure 3 shows that the combination of *CLN3* and *LRP10* reduced the within-group variation substantially compared with the expression ratios of the other combinations of reference genes. Furthermore, *CLN3* and *LRP10* showed lower variation among the medians in the different groups, suggesting that the different conditions had very small effects on the expression of these genes. The full figure showing all possible combinations of reference genes is shown in supplemental data (data available online at http://www.obesityresearch.org).

Discussion

In this study we used microarray data, real-time RT-PCR, and a bootstrap procedure to identify genes with low variation in expression levels in human adipose tissue. *LRP10* was identified as the gene with the highest expression stability in human adipose tissue biopsies that were selected to represent a wide range of commonly studied physiological parameters.

Several approaches have been made to adjust for sample variation in quantitative RT-PCR analysis. For example, gene expression has been related to total RNA or cell count (9). However, this approach lacks control of the cDNA synthesis step, and cell counts can only be applied to studies performed on cells and not tissues. Other strategies, such as spiking exogenous in vitro synthesized RNA (10,11), are time-consuming, which is why the majority of published studies use an endogenous reference gene. However, the evaluation of an optimal reference gene based on the real-time RT-PCR data becomes a circular problem because of the lack of an absolute reference point, as discussed by Vandesompele et al. (7) in their paper presenting a strategy to identify the most stably expressed reference genes (7). The use of the bootstrap technique, which we have applied in this study, extends the method of Vandesompele et al. by enabling us to differentiate the expression stability of the two highest ranked genes.

ACTB, 18S, and *GAPD* are the most frequently used reference genes for expression studies of human adipose

tissue, but other reference genes have also been used (e.g., *PPIA* and *RPLP0*). We found that of the established reference genes, *RPLP0* was highest ranked. However, expression of *RPLP0* when related to *LRP10* was significantly affected by diet-induced weight loss. *GAPD* expression in adipose tissue generally showed a higher variation in all groups compared with *18S* and *ACTB*, which is in contrast to in vitro cultured human adipocytes, where *GAPD* is stably expressed (12).

Little is known of the biological function of *LRP10*. The mouse homologue to *LRP10*, *Lrp10,* also known as *Lrp9*, mediates cellular uptake and hydrolysis of cholesterol esters in apolipoprotein E-enriched very-low-density lipoproteins in vitro (13). In conclusion, based on the results of this study, we recommend the use of *LRP10* as a reference gene for expression studies of human adipose tissue.

Acknowledgments

The Swedish Medical Research Council (11285, 14735, 14736), IngaBritt and Arne Lundberg Forskningsstiftelse, the Tore Nilsons Foundation for Medical Research, and the Fredrik and Ingrid Thurings Foundation supported this work. We thank Lars Sjöström, Jarl Torgerson, Hans ‚Lönroth, and Torsten Olbers for providing adipose tissue biopsies and Björn Carlsson for critically evaluating the manuscript.

References

1. **Schmittgen TD, Zakrajsek BA.** Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *J Biochem Biophys Methods*. 2000;46:69–81.
2. **Suzuki T, Higgins PJ, Crawford DR.** Control selection for RNA quantitation. *BioTechniques*. 2000;29:332–7.
3. **Bustin SA.** Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol*. 2002;29:23–39.
4. **Gabrielsson BG, Karlsson AC, Lönn M, et al.** Molecular characterization of a local of sulfonylurea system in human adipose tissue. *Mol Cell Biochem*. 2004;258:65–71.
5. **Gabrielsson BG, Johansson JM, Jennische E, et al.** Depot-specific expression of fibroblast growth factors in human adipose tissue. *Obes Res*. 2002;10:608–16.
6. **Gabrielsson BG, Johansson JM, Lonn M, et al.** High expression of complement components in omental adipose tissue in obese men. *Obes Res*. 2003;11:699–708.
7. **Vandesompele J, De Preter K, Pattyn F, et al.** 2002 Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 3:0034.
8. **Boos DB.** Introduction to the bootstrap world. *Stat Sci*. 2003; 18:168–74.
9. **Tricarico C, Pinzani P, Bianchi S, et al.** Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. *Anal Biochem*. 2002;309:293–300.
10. **Smith RD, Brown B, Ikonomi P, Schechter AN.** Exogenous reference RNA for normalization of real-time quantitative PCR. *Biotechniques*. 2003;34:88–91.
11. **Crawford EL, Peters GJ, Noordhuis P, et al.** Reproducible gene expression measurement among multiple laboratories obtained in a blinded study using standardized RT (StaRT)-PCR*1. *Molec Diagn*. 2001;6:217–25.
12. **Gorzelniak K, Janke J, Engeli S, Sharma AM.** Validation of endogenous controls for gene expression studies in human adipocytes and preadipocytes. *Horm Metab Res*. 2001;33: 625–7.
13. **Sugiyama T, Kumagai H, Morikawa Y, et al.** A novel low-density lipoprotein receptor-related protein mediating cellular uptake of apolipoprotein E-enriched beta-VLDL in vitro. *Biochemistry*. 2000;39:15817–25.

Supplemental data

Methods and Procedures

Subjects and samples

The characteristics of the subjects in the study are shown in Table S1 and S2.

RNA isolation and hybridization to the microarrays

Briefly, total RNA was isolated from the biopsies using the Chomczynski method (1) but increasing the ratio of chloroform/phenol to 1:1. Synthesis of biotin-labeled cRNA from total RNA (5 to 8 µg) and hybridization to the DNA microarrays were performed according to Affymetrix GeneChip® Expression Analysis manual (Affymetrix, Santa Clara, CA, USA). The microarrays were scanned with a confocal laser scanner and visualized using GeneChip® 4.0 software (Affymetrix). The average target signal on each microarray was globally scaled to an average intensity of 500. The scanned images were analyzed using Microarray Suite Version 5.0 software (Affymetrix).

Selection of stably expressed genes in human adipose tissue using microarray data

In total, 52 expression profiles of human adipose tissue were used for the selection (see Table S1). A pre-selection based on within group variation was performed using the largest group of arrays, i.e. paired samples from subcutaneous and omental adipose tissue (2, 3). The 50 genes with the smallest coefficient of variation (CV, %SD) were selected. **The initial selection was performed on individual probe-sets. When the corresponding genes were identified we also searched for other probe-sets representing the same gene. Subsequently, the geometric mean and CV of each probe-set representing one gene, was calculated.**

Subsequently, to perform a selection based on physiological parameters of interest, eight groups were created as described below and in Table S3:

1. Obese men, omental depot
2. Obese men, subcutaneous depot
3. Obese women, before weight-loss
4. Obese women, during weight-loss
5. Obese women, after weight-loss
6. Obese women, with type 2 diabetes
7. Obese women, non-diabetic
8. Normal-weight women, post-menopausal

Microarrays #37-48 (Table S1) contained two parameters of interest, presence of diabetes and effect of weight-loss, and were therefore included in two groups in the selection (Table S3). The average expression for each group was calculated, followed by computation of the CVs for the 50 genes across the eight groups. The ten genes with the lowest CV were selected for subsequent studies (table S4). The expression stability of the ten putative and five established reference genes was evaluated by real-time RT-PCR analysis on a different set of samples.

Real-time RT-PCR

Probe-primer sets for the four genes; COBRA1, ENTPD6, PDAP1 and HDAC5 were designed using the Primer Express® software v2.0 (Applied Biosystems, Foster City, CA) ensuring that the amplicons spanned an exon junction to avoid amplification of genomic DNA. The sequences were as follows; COBRA1 (probe; 5'-FAM-TCA CCA GGT TCC TCC CGA TGC TCA-TAMRA-3', forward primer; 5'-GGA GCC CAA GAT GGA GGT AGA-3', reverse primer; 5'-TGT AGT CAT CCA CCA GGA AGG A-3'), ENTPD6 (probe; 5'-FAM-CCC AGG AGC AAA GTG CTG AAG CTCA-TAMRA-3', forward primer; 5'-CTC ACC TAC GTC AGC CTG CTA CT-3', reverse primer; 5'-AGC TGG TCT CAA CAT TGT CAA TTT T-3'), PDAP1 (probe; 5'-FAM-TCA TCT TCT TCA TCC TCA CTC TCA TCT GA -TAMRA-3', forward primer; 5'-AGG AGC AAA AAG AAG GTG GAG AT-3', reverse primer; 5'- CCT TTG CGC TTT TGC TGG TA-3') and HDAC5 (probe; 5'-FAM-TGA TGC CCA TTG CCC ACG AGT TCT-TAMRA-3', forward primer; 5'-GAG TAC CTT ACA GCC TTC AGG ACA GT, reverse primer; 5'-GCG GAG ACT AGG ACC ACA TCA). The remaining genes were analyzed using Applied Biosystems pre-designed assays on demand (Table S4).

Working standards were prepared from a large pool of different adipose tissue RNAs and standard cDNA was synthesized in parallel with the sample cDNAs. A standard curve, obtained by serial dilution of the standard cDNA (range 0.625 ng to 40 ng original RNA per well), was included on each plate. Amplification and detection of specific products were performed with the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems), using default cycle parameters. All samples were analyzed individually in triplicate. The background was manually set for each gene from cycle 3 to about 2 cycles before the signal increased using the linear amplification plot view. Threshold values were adjusted to approximately the inflection point in the logarithmic amplification plot view.

Ranking the putative reference genes

We performed a bootstrap step to evaluate the certainty of the ranking. The ranking method (4) was bootstrapped (5) by re-sampling from the original set of 44 sample files, where each file contained the expression values of the analyzed genes for one sample, and each time acquiring a random selection of 44 files. In each resample, one sample file could hence be included 0 to 44 times. The rank of each gene was subsequently calculated as described in the article under "Ranking the putative reference genes". The re-sampling procedure was repeated 10,000 times. The rank distribution of the 13 genes in the 10,000 samples gives an evaluation of the certainty of the ranks.

Testing for effects of physiological variables

To test for the effects of physiological parameters specified by groups a)-k) in Figure 3 and Figure S2 we performed t-tests and linear regression tests on log-ratios using LRP10 as reference gene.

Results

The Bland-Altman plot was used to investigate whether there was any systematic covariance between two reference genes that would not be detected by simple linear regression plots (6). Figure S1 shows Bland-Altman plots of the novel reference genes CLN3 versus LRP10 and of 18S versus GAPD. The 95% confidence intervals for the expression ratios of CLN3/LRP10 and 18S/GAPD were 0.58 to 1.35 and 0.19 to 2.0, respectively.

Effects of gender, BMI, age, depot and biopsy procedure

Figure S2 shows the expression ratios of different combinations of reference genes. The expression ratios of the two best ranked genes are shown in the bottom row for comparison. Relative expression of the five established reference genes related to LRP10, shown in the fifth column, resulted in few outlying medians. The different physiological parameters appeared to affect the expression ratios of the established reference genes to a larger extent when related to each other (shown in the first four columns) compared with when related to LRP10. Of the variables, BMI and diet-induced weight-loss appeared to introduce the largest systematic variations. The expression of the majority of the established reference genes when related to LRP10 indicated effects of diet-induced weight-loss but, possibly due to the small number of patients, most of the results were not significant. The only clear effect was the increase of RPLP0 with time during very low calorie diet (Figure S2, column LRP10 and row RPLP0, p=0.006 in linear regression test).

Supplemental figures

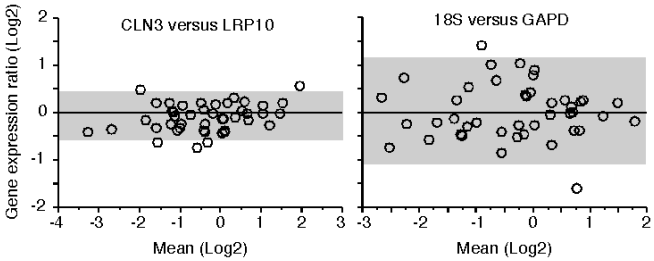


Figure S1. Bland-Altman plots of CLN3 versus LRP10 (left) and 18S versus GAPD (right). The gene expression was normalized for each gene before calculation of mean expression and expression ratio of the two reference genes in each sample (n=44). The mean of the ratios and the 95% confidence intervals are shown as a solid line and gray box, respectively.

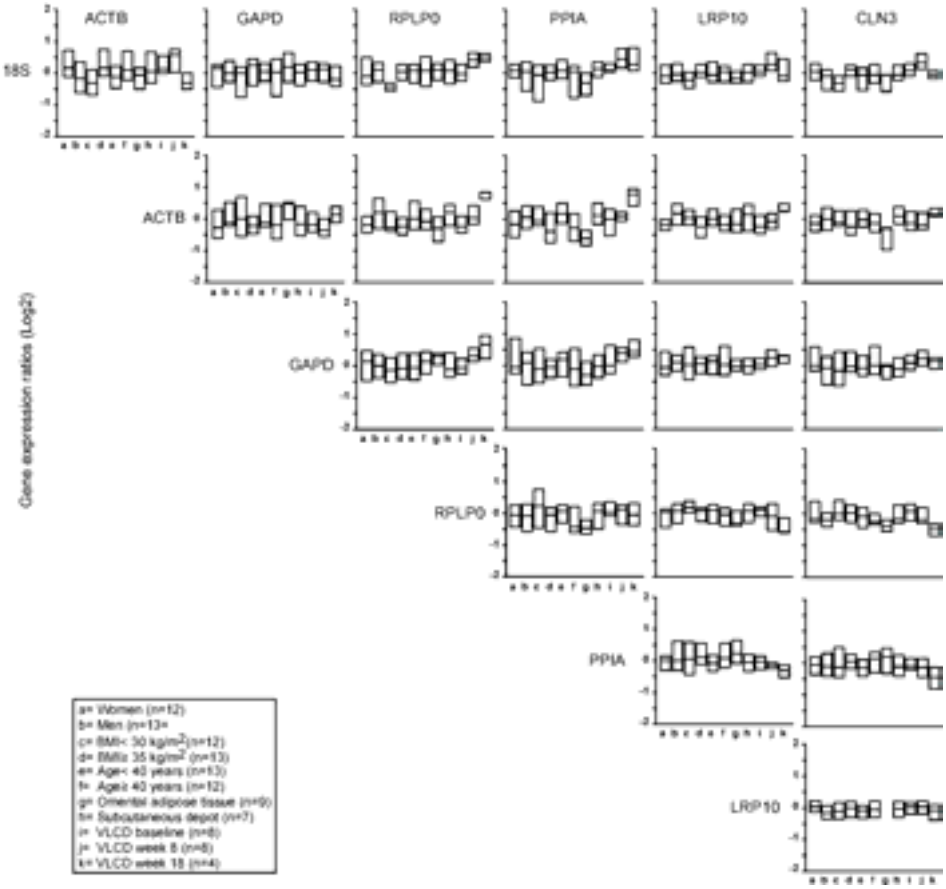


Figure S2: Variation in expression ratios of the different reference gene combinations in human adipose tissue analyzed by real-time RT-PCR The top row shows relative gene expression of ACTB/18S, GAPD/18S, RPLP0/18S etc; the bottom row shows relative gene expression of CLN3/LRP10, the two top ranked genes. The box plots show the quartiles of gene expression ratios where the expression of each gene was normalized before calculation of the ratios.

References

1. Chomczynski P, Sacchi N 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate- phenol-chloroform extraction. Anal. Biochem. 162:156-159.
2. Gabrielsson BG, Johansson JM, Jennische E et al. 2002 Depot-specific expression of fibroblast growth factors in human adipose tissue. Obes Res 10:608-16.

3. Gabrielsson BG, Johansson JM, Lönn M et al. 2003 High expression of complement components in omental adipose tissue in obese men. Obes Res 11:699-708.
4. Vandesompele J, De Preter K, Pattyn F et al. 2002 Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:RESEARCH0034
5. Boos DB 2003 Introduction to the bootstrap world. Statistical Science 18:168-174.
6. Bland JM, Altman DG 1986 Statistical methods for assessing agreement between two methods of clinical measurement. Lancet: i307-310.

Tables

Table S1. Characteristics of subjects included in the microarray studies.

Subject	Array #	Age (years)	Sex	Biopsy	Depot	Anesthesia	BMI (kg/m2)	S-Insulin (mU/l)
1	1, 2, 3, 4, 5, 6	47	male	surgical	omental and subcutaneous	general	41.2	37
2	7, 8, 9, 10, 11, 12	33	male	surgical	omental and subcutaneous	general	51.2	32
3	13, 14, 15, 16, 17, 18	51	male	surgical	omental and subcutaneous	general	39.4	12
4	19, 20, 21, 22, 23, 24	37	male	surgical	omental and subcutaneous	general	41	19
5	25, 26, 27, 28, 29,30	41	male	surgical	omental and subcutaneous	general	42.3	8
6	31, 32,33, 34, 35, 36	56	male	surgical	omental and subcutaneous	general	39	38
7	37, 38, 39, 40, 41, 42	50	female	needle	subcutaneous	local	31.4	24
8	37, 38, 39, 40, 41, 42	59	female	needle	subcutaneous	local	37.5	43
9	37, 38, 39, 40, 41, 42	48	female	needle	subcutaneous	local	39.0	16
10	37, 38, 39, 40, 41, 42	47	female	needle	subcutaneous	local	38.5	15
11	37, 38, 39, 40, 41, 42	41	female	needle	subcutaneous	local	39.2	16
12	37, 38, 39, 40, 41, 42	57	female	needle	subcutaneous	local	38.5	20
13	37, 38, 39, 40, 41, 42	51	female	needle	subcutaneous	local	45.9	21
14	37, 38, 39, 40, 41, 42	59	female	needle	subcutaneous	local	46.9	34
15	37, 38, 39, 40, 41, 42	35	female	needle	subcutaneous	local	59.0	70
16	43, 44, 45, 46, 47, 48	48	female	needle	subcutaneous	local	34.0	7
17	43, 44, 45, 46, 47, 48	56	female	needle	subcutaneous	local	35.3	12
18	43, 44, 45, 46, 47, 48	50	female	needle	subcutaneous	local	37.6	10
19	43, 44, 45, 46, 47, 48	49	female	needle	subcutaneous	local	36.8	9
20	43, 44, 45, 46, 47, 48	50	female	needle	subcutaneous	local	36.8	16
21	43, 44, 45, 46, 47, 48	51	female	needle	subcutaneous	local	39.7	12
22	43, 44, 45, 46, 47, 48	47	female	needle	subcutaneous	local	41.3	7
23	43, 44, 45, 46, 47, 48	37	female	needle	subcutaneous	local	43.3	14
24	43, 44, 45, 46, 47, 48	55	female	needle	subcutaneous	local	55.5	14
25	49, 50, 51, 52	58	female	needle	subcutaneous	local	26.9	54
26	49, 50, 51, 52	62	female	needle	subcutaneous	local	26.1	32
27	49, 50, 51, 52	60	female	needle	subcutaneous	local	31.5	46

Table S2. Characteristics of samples included in the real time RT-PCR study.

Subject	Sample	Age (years)	Sex	Type of biopsy	Depot	Anesthesia	BMI (kg/m2)	S-Insulin (mU/l)
1	1	33	male	surgical	omental	general	51.2	32
1	2	33	male	surgical	subcutaneous	general	51.2	32
2	3	51	male	surgical	omental	general	39.4	12
2	4	51	male	surgical	subcutaneous	general	39.4	12
3	5	40	male	surgical	omental	general	52.2	14
3	6	40	male	surgical	subcutaneous	general	52.2	14
4	7	31	male	surgical	omental	general	46.0	25
5	8	26	female	surgical	omental	general	61.2	46[1]
5	9	26	female	surgical	subcutaneous	general	61.2	46[1]
6	10	48	female	surgical	omental	general	48.1	9.2[1]
7	11	64	female	surgical	omental	general	36.5	67[1]
7	12	64	female	surgical	subcutaneous	general	36.5	67[1]
8	13	43	male	surgical	subcutaneous	local	26.0	5.6[1]
9	14	25	male	surgical	subcutaneous	local	24.7	7.6[1]
10	15	22	female	surgical	subcutaneous	local	26.3	9.2[1]
11	16	49	female	surgical	subcutaneous	local	24.0	5.4[1]
12	17	54	male	needle	subcutaneous	local	23.0	5.2[1]
13	18	33	male	needle	subcutaneous	local	24.9	5.0[1]
14	19	33	male	needle	subcutaneous	local	28.8	13[1]
15	20	42	male	needle	subcutaneous	local	20.5	4.7[1]
16	21	29	male	needle	subcutaneous	local	37.8	31
16	22	29	male	needle	subcutaneous	local	32.0	32
16	23	29	male	needle	subcutaneous	local	28.6	7
17	24	34	male	needle	subcutaneous	local	40.1	16
17	25	34	male	needle	subcutaneous	local	35.7	10
17	26	34	male	needle	subcutaneous	local	33.7	11
18	27	29	male	needle	subcutaneous	local	32.5	20
18	28	29	male	needle	subcutaneous	local	28.8	11
18	29	29	male	needle	subcutaneous	local	29.1	12
19	30	25	male	needle	subcutaneous	local	41.4	30
19	31	25	male	needle	subcutaneous	local	33.8	9
19	32	25	male	needle	subcutaneous	local	29.1	10
20	33	49	female	needle	subcutaneous	local	36.8	9
20	34	49	female	needle	subcutaneous	local	33.3	6
21	35	35	female	needle	subcutaneous	local	59.0	70
21	36	35	female	needle	subcutaneous	local	51.0	28
22	37	57	female	needle	subcutaneous	local	38.5	20
22	38	57	female	needle	subcutaneous	local	34.5	20
23	39	51	female	needle	subcutaneous	local	45.8	21
23	40	51	female	needle	subcutaneous	local	39.6	10
24	41	46	female	needle	subcutaneous	local	26.9	4.3
25	42	20	female	needle	subcutaneous	local	29.0	19
26	43	54	female	needle	subcutaneous	local	25.0	5.6
27	44	46	female	needle	subcutaneous	local	26.1	6.6

[1] Plasma insulin

Table S3. Grouping the expression profiles into the eight physiological parameters of interest.

Group	Array #	Samples
1. Obese men, omental depot	1, 2, 3, 7, 8, 9, 13, 14, 15, 19, 20, 21, 25, 26, 27, 31, 32, 33	Individual RNA from omental adipose tissue (surgical) from 6 obese men (triplicate arrays)
2. Obese men, subcutaneous depot	4, 5, 6, 10, 11, 12, 16, 17, 18, 22, 23, 24, 28, 29, 30, 34, 35, 36	Individual RNA from subcutaneous adipose tissue (surgical) from 6 obese men (triplicate arrays)
3. Obese women, before weight-loss	37, 38, 43, 44	RNA pooled from subcutaneous adipose tissue from 18 obese women during weight-loss; before start of a very low calorie diet (VLCD, duplicate arrays)
4. Obese women, during weight-loss	39, 40, 45, 46	RNA pooled from subcutaneous adipose tissue from 18 obese women during weight-loss; after 8 weeks of VLCD (duplicate arrays)
5. Obese women, after weight-loss	41, 42, 47, 48	RNA pooled from subcutaneous adipose tissue from 18 obese women during weight-loss; two weeks after completed 16-week VLCD (duplicate arrays)
6. Obese women, with type 2 diabetes	37, 38, 39, 40, 41, 42	RNA pooled from subcutaneous adipose tissue from 9 obese women with type 2 diabetes (duplicate arrays)
7. Obese women, non-diabetics	43, 44, 45, 46, 47, 48	RNA pooled from subcutaneous adipose tissue from 9 obese women with no diabetes (duplicate arrays)
8. Normal-weight women, post-menopausal	49, 50, 51, 52	RNA pooled from subcutaneous adipose tissue from 3 post-menopausal women before and after seven days daily injection of 25 mg prednisolone (duplicate arrays)

Table S4. Novel and established reference genes for expression analysis in human adipose tissue

Gene symbol	Signal (mean±SD)*	CV (%)*	Gene name	Function	ABI Assay on demand
LRP10	1861 ± 62	3.4	Low density lipoprotein receptor-related protein 10	Unknown (cholesterol metabolism)	Hs00204094_m1
HPCAL1	3977 ± 165	4.3	Hippocalcin-like 1	Unknown, (calcium-BP)	Hs00365962_m1
COBRA1	3120 ± 146	4.7	Cofactor of BRCA1	Chromatin structure, transcription regulation	Own probe-primer design
HDAC5	1054 ± 62	5.9	Histone deacetylase 5	Chromatin modeling/ silencing	Own probe-primer design
ENTPD6	1206 ± 77	6.4	Ectonucleoside triphosphate diphosphohydrolase 6	Unknown, member of the CD39-like family	Own probe-primer design
NME3	1172 ± 86	7.3	Protein expressed in non-metastatic cells 3	Nucleotide kinase, pyrimidine metabolism	Hs00358004_g1
PDAP1	2835 ± 212	7.5	PDGFA associated protein 1	Unknown	Own probe-primer design
PSAP	17360 ±1344	7.7	Prosaposin (sphingolipid activator protein-1)	Lipid metabolism/ transport/binding	Hs00248055_m1
CLN3	1099 ± 91	8.2	Ceroid-lipofuscinosis, neuronal 3, juvenile	Putative mitochondrial membrane protein; protein folding/ chaperone	Hs00164002_m1
MGAT1	2354 ± 201	8.5	Mannosyl (alpha-1,3)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase	Protein glycosylation, carbohydrate metabolism	Hs00159121_m1
ACTB	17802 ± 1315 23378 ± 4514 13236 ± 2755 22745 ± 6627	7.4 19.3 20.8 29.1	Beta actin	Cytoskeletal structure (nonmuscle)	Hs99999903_m1
B2M	15415 ± 10623 13248 ± 2392 17082 ± 2539	68.9 18.1 14.9	Beta-2-microglobulin	Immune response	Not assayed
GAPD	14164 ± 2290 7267 ± 1306 10595 ± 2834 17340 ± 6987	16.2 18.1 26.7 40.3	Glyceraldehyde-3-phosphate dehydrogenase	Carbohydrate metabolism	Hs99999905_m1
GUSB	2305 ± 619	26.9	Glucuronidase, beta	Glycosaminoglycan catabolism	Not assayed
PPIA	15696 ± 1578	10.1	Peptidylprolyl isomerase A (cyclophilin A)	Cyclosporin A binding protein	Hs99999904_m1
RPLP0	12572 ± 1626	12.9	Large ribosomal protein P0	Ribosomal protein	Hs99999902_m1
SDHA	2665 ± 341	12.8	Succinate dehydrogenase complex, subunit A	Complex II of the respiratory chain	Not assayed
TBP	205 ± 54	33.3	TATA-box binding protein	Transcription by RNA polymerase I, II and III	Not assayed
UBC	21091 ± 4274 6008 ± 2786	20.3 46.4	Ubiquitin C	Protein degradation	Not assayed
YWHAZ	1479 ± 498	33.7	Phospholipase A2	Glycerolipid and phospholipid metabolism	Not assayed
18S	NA	NA	18S ribosomal RNA		Hs99999901_s1

* Signal intensities from microarray data (arbitary units).
NA: not applicable

Paper III

Weighted analysis of paired microarray experiments

Erik Kristiansson*, Anders Sjögren*†,
Mats Rudemo and Olle Nerman

Abstract

In microarray experiments quality often varies, for example between samples and between arrays. The need for quality control is therefore strong. A statistical model and a corresponding analysis method is suggested for experiments with pairing, including designs with individuals observed before and after treatment and many experiments with two-colour spotted arrays. The model is of mixed type with some parameters estimated by an empirical Bayes method. Differences in quality are modelled by individual variances and correlations between repetitions. The method is applied to three real and several simulated datasets. Two of the real datasets are of Affymetrix type with patients profiled before and after treatment, and the third dataset is of two-colour spotted cDNA type. In all cases, the patients or arrays had different estimated variances, leading to distinctly unequal weights in the analysis. For simulated data the improvement relative to previously published methods without weighting is shown to be substantial.

Keywords

Quality control, QC, Quality Assurance, QA, Quality Assessment, Empirical Bayes, DNA Microarray

*both authors contributed equally, order was randomised
†corresponding author; e-mail: anders.sjogren@math.chalmers.se

1 Introduction

DNA microarrays are strikingly efficient tools for analysing gene expression for large sets of genes simultaneously. They are often used to identify genes that are differentially expressed between two conditions, e.g. before and after some treatment. A drawback is that the technology involves several consecutive steps, each exhibiting large quality variation. Thus there is a strong need for quality assessment and quality control to handle occurrences of poor quality, as is clearly pointed out in Johnson and Lin (2003) and Shi et al. (2004).

Despite the observed need for effective quality control, only recently have standard operating procedures for quality assurance of the entire chain of processing steps been proposed (Ryan et al., 2004, for one-channel experiments only). However, even utilising an optimal quality control procedure aiming at removing low quality arrays and/or individual gene measurements (e.g. spots), there will always be a marginal region with some measurements being of decreased quality without being worthless, as noted in Ryan et al. (2004). Consequently, it should be possible to make progress by integrating quality control quantitatively into the analysis following the lab steps and low-level analysis, taking quality variations into account.

When integrating the quality concept into the analysis, the quality of different parts of the dataset should ideally be estimated from data and used in the subsequent selection of differentially expressed genes. Here we introduce a method, called *Weighted Analysis of paired Microarray Experiments* (referred to as WAME), for the analysis of paired microarray experiments, e.g. comparison of pairs of treatment conditions and most two-colour experiments. WAME aims at estimating array-wide quality deviations and integrates the quality estimates in the statistical analysis. Only the observed gene expression ratios are used in the quality assessment, making the method applicable to most paired microarray experiments, independent of which DNA microarray technology is used.

In short WAME identifies and downweights repetitions (biological or technical) of pairs (corresponding to individuals or to arrays) with decreased quality for many genes. Repetitions with positively correlated variations, e.g. caused by shared sources of variation, are similarly down-weighted. Thus, estimates of differential expression with improved precision and tests with increased power are provided.

In the adopted model, log ratios of measured RNA-levels are assumed normally distributed. The covariance structure is specified by parameters of two types: (i) a global covariance matrix signifying different quality for dif-

ferent repetitions and (ii) gene specific multiplicative factors. The latter have inverse gamma prior distribution with one gene-specific parameter, which is estimated by an empirical Bayes method.

The paper is organised as follows. In the next section, a background and a selection of previous work in the field are presented. This is followed by a detailed description of the our model. Methods for estimating the parameters and a likelihood ratio test for identifying differentially expressed genes are derived. In the following section simulations are used to compare WAME to four currently used methods: (i) average fold change, (ii) ordinary t-test, (iii) the penalized t-statistic of Efron et al. (2001), and (iv) the moderated t-statistic of Smyth (2004). Next, WAME is applied to three real datasets, the *Cardiac* dataset of Hall et al. (2004), the *Polyp* dataset of Benson et al. (2004) and the *Swirl* dataset (Dudoit and Yang, 2003). The results obtained are discussed in a subsequent section and some derivations and mathematical details are given in an appendix.

2 Background

To put the quality control aspect of our model into context, the different steps and sources of variation in typical paired microarray experiments are outlined below. In addition, a selection of publications dealing with quality control for microarray experiments are briefly reviewed.

2.1 Sources of variation in typical microarray experiments

The first step, after decision on experimental design, of a microarray experiment aiming at identifying differentially expressed genes would typically be to determine how biological samples should be acquired. In experiments dealing with homogeneous groups of single cell organisms, such as yeast, in highly controlled environments, this task is typically less complex than when dealing with heterogeneous groups of multicellular organisms, such as humans. Here selection of subjects and cells from the relevant organ, e.g. by biopsy or laser dissection, are complicated tasks.

From the biological sample the following lab-steps are performed: RNA extraction, reverse transcription (and *in vitro* transcription), labelling, hybridisation to arrays and scanning. The parts of the scanned images corresponding to the different genes (i.e. spots or probe-pairs) are identified and quantified. In addition, background correction may be performed. Subse-

quently, normalisation of the quantified measurements is performed to handle global differences. In the case of Affymetrix type arrays, 11-20 pairs of quantitative measurements are combined into one expression level estimate for each gene. For an experiment of paired type, one $\log_2$-ratio of the expression level estimates is calculated for each pair and gene. These $\log_2$-ratios are then used to examine which genes are differentially expressed.

In several of the steps mentioned above there are substantial quality variations. For example, the quantity and quality of the RNA in biopsies may vary considerably. There are sometimes evidence of poor quality making it possible to remove obviously worthless samples. Nevertheless, there will always be a marginal region with some measurements being of reduced quality without being worthless, as noted in Ryan et al. (2004). In addition, some variations are hard to detect before the actual normalised $\log_2$-ratios are computed, e.g. the representativeness in tissue distribution of human biopsies. An additional aspect of quality control is systematic errors, where the variations of different repetitions are correlated. This could be due to shared sources of variation, such as simultaneous processing in lab steps or similar, non-representative tissue composition in the biopsies.

Another potentially important factor is the quality of the arrays used for the measurements. Flaws in the manufacturing process might make measurements for individual genes inferior. This is more of a problem in the case of spotted arrays, since there is only one or a few spots per gene. However, such bad spots can often be detected. The quality control in the actual manufacturing of microarrays is certainly very important but will not be further discussed here.

2.2 A brief review of some relevant publications

In Johnson and Lin (2003) and Shi et al. (2004) the general need for improved quality assurance in the context of DNA microarray analysis is emphasised. Tong et al. (2004) implement a public microarray data and analysis software and note that "Although the importance of quality control (QC) is generally understood, there is little QC practise in the existing microarray databases". They include some available measures of quality for different steps in the analysis in their database.

Dumur et al. (2004) survey quality control criteria for the wet lab steps of Affymetrix arrays, going from RNA to cDNA. Additionally, three sources of technical variation (hybridisation day, fluidic scan station, fresh or frozen cDNA) are evaluated using an ANOVA model.

Ryan et al. (2004) present guidelines for quality assurance of Affymetrix based microarray studies, utilising a variety of techniques for the different steps, some of which are shown to agree. A sample quality control flow diagram is suggested, including steps from extracted RNA to the quantified arrays.

Chen (2004) aims at screening out ineligible arrays (Affymetrix type), using a graphical approach, so called *2D image plots*, to display grouped data. Park et al. (2005) similarly aim at identifying outlying slides in two-channel experiments by using scatterplots of transformed versions of the signals from the two channels.

Tomita et al. (2004) use correlation between arrays (Affymetrix type) to evaluate the RNA integrity of the individual arrays, by forming an average correlation index (ACI). The ACI is shown to correlate with several existing quality factors, such as the 3'-5' ratio of GAPDH.

Several papers have been written on the quality control of individual measurements of genes (spots or probes). Wang et al. (2001, 2003) define a spot-wise composite score from various quantitative measures of quality of individual spots in spotted microarrays. They further perform evaluations on several in-house datasets, showing that when bad spots are removed, the variance of all gene-wise ratios in one chip is decreased. In Hautaniemi et al. (2003) Bayesian networks are used to discriminate between good and bad spots with training data provided by letting experienced microarray users examine the arrays by hand.

In the papers discussed above the countermeasure against low-quality spots or arrays is to treat them as outliers and to remove them. Again, there will always be a marginal region with some measurements being of decreased quality without being worthless. An interesting approach using weighted analysis of the microarray gene expression data is due to Bakewell and Wit (2005). The starting point is a variance component model for the log expression mean for a spot i with variance $\sigma_b^2 + \sigma_i^2/m_i$, where σ_b^2 is the variance between spots while σ_i^2 is the variance between pixels in spot i with the effective number m_i of pixels. For each gene the spots are weighted inversely proportional to estimated variances, and different genes are essentially treated independent of each other. Only quality deviations of the actual hybridised spots are included in the model.

In Yang et al. (2002) the variance of different print tip groups or arrays in cDNA experiments are estimated by a robust method. The need for scale normalisation between slides is determined empirically, e.g. by displaying box plots for the log ratios of the slides.

The model we propose (WAME) assesses the quality of different arrays

quantitatively by examining the computed $\log_2$-ratios. Thus, quality deviations in all steps leading to the gene expression estimates are included, as long as the quality deviations occur for a wide variety of measured genes. Furthermore, shared systematic errors are taken care of via estimated covariances between repetitions. The assessed qualities are incorporated into the analysis based on the statistical model presented in the next sections.

In microarray experiments there are often relatively few replicates, resulting in highly variable gene-wise variance estimates. To use the information in the large number of measured genes to handle this problem, an empirical Bayes approach can be taken, determining a prior distribution from the data, thus moderating extreme estimates. This approach has been used in Baldi and Long (2001), Lönnstedt and Speed (2002) and Smyth (2004).

3 The model

The experimental layouts studied in the present paper are oneswith comparison of paired observations. For each gene $g = 1, \ldots, N_G$ and each pair of measurements $i = 1, \ldots, N_I$, let X_{gi} be the normalised $\log_2$-ratio of the gene expressions from the two conditions. The expected value μ_g of X_{gi} measures the log ratio of the RNA concentrations.

In Section 2.1 it was noted that there may exist dependencies between repetitions, e.g. due to systematic errors. Furthermore, different arrays may have different precision in their measurements of the gene expressions. To describe this, we use a covariance structure matrix Σ which models precision as individual variances for the different repetitions and dependencies between repetitions as covariances.

Due to both technical and biological reasons the observations for the different genes have different variability, and a gene-specific multiplicative factor c_g for the covariance matrix is introduced. The c_g-variables for different genes are assumed to be independent. Given c_g the vector $\mathbf{X_g}$ consisting of all repetitions for gene g is assumed to have a N_I-dimensional normal distribution with mean vector $\mu_g\mathbf{1}$ and covariance matrix $c_g\Sigma$. The vectors $\mathbf{X_g}$ for different genes are also assumed independent.

In microarray experiments, the number of experimental units is typically fairly small and estimates of c_g utilising only information from the measurements with gene g would be highly variable. Therefore prior information is introduced as an a priori distribution for c_g, which serves to moderate the estimates of c_g. The prior for c_g is assumed to be an inverse gamma distribution with a parameter α determining the spread of the distribution, in

effect determining the information content in the prior. The inverse gamma distribution is a conjugate prior distribution for the variance of a normal distribution and has as such been used in Bayesian and empirical Bayesian analysis of microarray data before (Baldi and Long, 2001; Lönnstedt and Speed, 2002; Smyth, 2004).

The model can be summarised as follows: We observe $\mathbf{X}_g = (X_{g1}, \ldots, X_{gN_I})$ where $g = 1, \ldots, N_G$. Let Σ be a covariance matrix with N_I rows and columns, c_g a set of gene-wise variance scaling factors and α a hyperparameter determining the spread of the prior distribution for c_g. Then for fixed μ_g, Σ and α,

$$\begin{aligned} c_g &\sim \Gamma^{-1}(\alpha, 1) \quad \text{and} \\ \mathbf{X}_g \mid c_g &\sim \mathbf{N}_{N_I}\left(\mu_g \mathbf{1}, c_g \Sigma\right), \end{aligned} \tag{1}$$

and all variables corresponding to different genes are assumed independent.

4 Inference

4.1 Estimation of a scaled version of the matrix Σ

Estimating Σ may appear easy but it turns out to be rather intricate and there are several issues involved.

Firstly, there are trivial solutions that give infinite likelihood of the model. Just put one gene-specific mean value μ_g equal to the observation of one of the repetitions and the corresponding variance equal to zero. To avoid this complication the assumption that the differential expression of most genes is approximately zero is introduced temporarily. This assumption is not as consequential as it might sound, since it is made by most of the procedures that have become *de facto* standard in the (preceding) normalisation step, one example being the loess normalisation method (Yang et al., 2002). Nevertheless, it does limit the set of experimental setups that can be treated and the proportion of genes that are regulated must not be too large. The impact of this assumption is further investigated by a simulation study in Section 5.2. For the rest of this section, μ_g is thus set equal to zero for all $g = 1, \ldots, N_G$.

Another issue is the scaling of Σ. For each gene, the covariance matrix is scaled with the random variable c_g which has an inverse gamma distribution with a parameter which is unknown in a first stage. To address this issue, the estimation of Σ is performed in two steps. In the first step, a transformation is applied to $\mathbf{X}_g$ such that the transformed vector has a distribution that is

independent of c_g. To simplify notation the index g will be dropped from $\mathbf{X}_g$ and c_g in the rest of this section. Let $\mathbf{U} = (U_1, \ldots, U_{N_I})$ where

$$U_i = \begin{cases} X_1 & \text{if } i = 1 \\ X_i/X_1 & \text{if } 2 \leq i \leq N_I. \end{cases}$$

The distribution of the vector $\mathbf{U}$ has the density

$$f_{\mathbf{U} \mid c,\Sigma}(\mathbf{u}) = f_{\mathbf{X} \mid c,\Sigma}(\mathbf{x}(\mathbf{u}))|J(\mathbf{u})|$$

where J is the corresponding Jacobian. Some algebra shows that the scaling factor c cancels for $U_2, \ldots, U_{N_I}$ and by integrating over U_1, we get the density

$$\begin{aligned} f_{U_2,\ldots,U_{N_I} \mid \Sigma}(u_2, \ldots, u_{N_I}) &= \int_{-\infty}^{\infty} f_{\mathbf{U} \mid c,\Sigma}(\mathbf{u})\, du_1 \\ &= C\, |\Sigma|^{-1/2} \left[v^{\mathrm{T}} \Sigma^{-1} v\right]^{-N_I/2}, \end{aligned} \tag{2}$$

where C is a normalisation constant and $v = (1, u_2, \ldots, u_{N_I})$. The distribution (2) is independent of c and the marginal distributions are scaled and translated Cauchy distributions.

From (2) we see that the distribution of $\mathbf{U}$ is unchanged if we multiply Σ with a constant. Let us therefore fix one element of Σ, e.g. we put the first element in the first row equal to one. Let Σ^* denote the matrix thus obtained. Then

$$\Sigma^* = \lambda \Sigma\,, \tag{3}$$

and the constant λ will estimated together with the hyperparameter α as described below in Section 4.2. Thus estimation of the covariance matrix Σ will be carried out in two steps: first estimate Σ^* with one element fixed and then estimate λ.

Numerical maximum likelihood based on the distribution (2) is used to produce a point estimate of Σ^*. The computational complexity grows as N_I^2 since the number of unknown parameters $N_I(N_I + 1)/2$. To get an efficient implementation C/C++ is combined with R (R Development Core Team, 2004). The resulting computational time for three arrays is less than a second and for 30 arrays it takes a few hours.

4.2 Estimation of the hyperparameter α and the scale λ

In this section, we develop methods for estimation of the hyperparameter α as well as the scale parameter λ in (3). From the model assumptions

in Section 3 we recall that c_g has as an inverse gamma distribution with hyperparameter α, e.g.

$$c_g \mid \alpha \sim \Gamma^{-1}(\alpha, 1).$$

The inference of α will be based on the statistic

$$S_g = (A\mathbf{X}_g)(A\Sigma A^{\mathrm{T}})^{-1}A\mathbf{X}_g,$$

where A is an arbitrary contrast matrix, i.e. a $(N_I - 1) \times N_I$ matrix with full rank and each row sum equal to 0. It follows that the distribution of S_g conditioned on c_g is a scaled chi-squared distribution with $N_I - 1$ degrees of freedom,

$$S_g \mid c_g \sim c_g \cdot \chi^2_{N_I-1}.$$

The unconditional distribution of S_g can be calculated by use of the fact that a gamma distribution divided by another gamma distribution has an analytically known distribution, a beta prime distribution (Johnson et al., 1995). Thus,

$$S_g \mid \alpha \sim 2 \times \beta' \left((N_I - 1)/2, \alpha\right),$$

which has the density function

$$f_{S_g \mid \alpha}(s_g) = \frac{1}{2}\frac{\Gamma(\alpha + (N_I - 1)/2)}{\Gamma(\alpha)\Gamma((N_I - 1)/2)}\frac{(s_g/2)^{(N_I-1)/2-1}}{[1 + s_g/2]^{\alpha+(N_I-1)/2}}.$$

In the same fashion, denote the variance estimator based on Σ^* in (3) by S^*_g, that is,

$$S^*_g = (AX_g)^{\mathrm{T}}(A^{\mathrm{T}}\Sigma^* A)^{-1}AX_g\ .$$

It follows that, $S^*_g = S_g/\lambda$ so

$$S^*_g \mid \alpha, \lambda \sim 2/\lambda \times \beta' \left((N_I - 1)/2, \alpha\right)\ .$$

Assuming independence between the genes, α and λ can now be estimated by numerical maximum likelihood. The estimated value of the (unscaled) covariance matrix Σ can then be calculated from Equation (3). Results from simulations show that the estimation of α and λ is accurate enough for any realistic values (results not shown). In the following sections, these parameters are assumed to be known.

4.3 The posterior distribution of c_g

The posteriori distribution of c_g is not explicitly used in the calculations above, but still of general interest. As previously mentioned, the distribution of S_g conditioned on c_g is a scaled chi-squared distribution with $N_I - 1$ degrees of freedom. Since chi-squared distributions and inverse gamma distributions are conjugates, the posteriori distribution of c_g given S_g is an inverse gamma distribution as well. We find

$$c_g \sim \Gamma^{-1}(\alpha, 1)$$

$$c_g \mid S_g \sim \Gamma^{-1}\left(\alpha + (N_I - 1)/2, 1 + \frac{S_g}{2}\right),$$

and the prior can be interpreted as representing 2α pseudo observations, which add a common variance estimate to all genes. A discussion regarding the use of this model in microarray analysis can be found in Lönnstedt and Speed (2002) and Smyth (2004) and a general discussion in Robert (2003).

4.4 Inference about μ_g

In this section we derive a statistical test for differential expression based on the WAME model. The hypotheses can for gene g can be formulated as

$$\begin{aligned} H_0 &: \text{ gene } g \text{ is not regulated } (\mu_g = 0) \\ H_A &: \text{ gene } g \text{ is regulated } (\mu_g \neq 0). \end{aligned}$$

A test suitable for the hypothesis H_0 is the likelihood ratio test (LRT) based on the ratio of the maximum values of the likelihood function under the different hypotheses. With our notation we reject H if

$$\frac{\sup\limits_{H_A} L(\mu_g|\mathbf{x}_g)}{\sup\limits_{H} L(\mu_g|\mathbf{x}_g)} = \frac{\sup\limits_{\mu_g \neq 0} L(\mu_g|\mathbf{x}_g)}{L(0|\mathbf{x}_g)} \geq k, \tag{4}$$

where k, $1 \leq k < \infty$, sets the level of the test. To calculate the likelihood function, we need to integrate over c_g, e.g.,

$$\begin{aligned} L(\mu_g|\mathbf{x}) &= \int f_{\mathbf{x}|\mu_g,c_g}(\mathbf{x}|c_g) f_{c_g|\alpha}(c_g|\alpha)\, dc_g \\ &= (2\pi)^{-N_I/2} |\Sigma|^{-1/2} \frac{\Gamma(N_I/2 + \alpha)}{\Gamma(\alpha)} \left[\frac{(\mathbf{x}_g - \mu_g \mathbf{1})^{\mathrm{T}} \Sigma^{-1} (\mathbf{x}_g - \mu_g \mathbf{1})}{2} + 1\right]^{-(\alpha + N_I/2)}. \end{aligned}$$

It is now straight forward to calculate the denominator $L(0|\mathbf{x}_g)$ in (4) and some algebra shows that the numerator is maximised by $\hat{\mu}_g = \bar{x}_g^w$, where

$$\bar{x}_g^w = \frac{\mathbf{1}^{\mathrm{T}}\Sigma^{-1}}{\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}}\mathbf{x}_g \ ,$$

is a *weighted mean value* of the observations. Analogously, we define the random variable $\bar{X}_g^w$ by replacing $\mathbf{x}_g$ with $\mathbf{X}_g$. It can be shown that

$$\mathbf{w}^{\mathrm{T}} = \frac{\mathbf{1}^{\mathrm{T}}\Sigma^{-1}}{\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}} \tag{5}$$

is the weight vector that minimises the variance of $\mathbf{w}^{\mathrm{T}}\mathbf{X}_g$. The weights in equation (5) will depend on the covariance matrix as follows. A repetition with high variance will have a low weight while a repetition with low variance will have a high weight. Moreover, a positive high correlation between repetitions will cause decreased weights. Note that if a repetition is highly correlated with a repetition with lower variance, its weight can actually become negative. According to the theory, this is nothing strange but practically this is of course not satisfying. Fortunately, such extreme cases seem to be rare in the microarray context and if they appear, the source of the correlation should be investigated and one could consider removing the negatively weighted repetition.

Evaluation of the likelihood function at 0 and $\bar{x}_g^w$ and a few calculations show that the inequality (4) is equivalent to

$$\frac{|\bar{x}_g^w|}{\sqrt{s_g + 2}} \geq k'$$

where s_g is defined according to Section 4.2 and k' is a new constant ($0 \leq k' < \infty$). Thus if we define the statistic T_g as

$$T_g = \frac{\bar{X}_g^w}{\sqrt{S_g + 2}}.$$

the null hypothesis is rejected if

$$|T_g| \geq k'.$$

The statistic T_g will be referred to as the *weighted moderated t-statistic* since it is a weighted generalisation of the moderated t-statistic derived by Lönnstedt and Speed (2002) and later refined by Smyth (2004). Indeed, if the weights

are equal (i.e. all weights are equal to $1/n$), T_g becomes equivalent to the result in Section 3 in Smyth (2004). To calculate the value of k' that corresponds to a given level of the test, the distribution of T_g needs to be derived. It turns out to be a scaled t-distribution with $2\alpha + N_I - 1$ degrees of freedom,

$$T_g \sim \frac{1}{\sqrt{2\alpha + N_I - 1}} \times t_{2\alpha+N_I-1}(\mu_g).$$

5 Results from simulations

5.1 Comparison to similar gene ranking methods

A simulation study was done to compare the performance of WAME to four published methods. These methods were

- Average fold-change
- Ordinary t-statistic
- Efron's penalized t-statistic
- Smyth's moderated t-statistic

The average fold-change for a gene is simply the mean value over all the observed $\log_2$-ratios and the ordinary t-statistic is the average fold-change divided by the corresponding sample standard deviation. These two methods have traditionally been popular gene ranking methods and it is therefore interesting to see how they perform. Another method introduced in (Efron et al., 2001) is the penalized t-statistic which is a modified version of the ordinary t-statistic where a constant has been added to the sample standard deviation. The motivation for this adjustment is the unreliability of the t-statistic in situations when only a few repetitions are used. The constant used here was chosen as the 90th percentile of the empirical distribution of the sample standard deviations, according to Efron et al. (2001). Finally, the moderated t-statistic is included. It was developed and implemented Smyth (2004) and it is available in the R package LIMMA (Smyth et al., 2003). The moderated t-statistic can be seen as a refined version of the B-statistic which was first presented in Lönnstedt and Speed (2002). In the paired microarray context, WAME is a generalisation of LIMMA in the sense that the two models are identical when all repetitions have the same variance and no correlations exist.

All methods were applied to a series of simulated datasets with different settings and the number of true positives as a function of false positives was plotted, generating several so called receiver operator characteristic (ROC) curves. The average over 100 datasets was used to produce a single curve where each dataset was created as follows. The number of genes (N_G) was fixed to 10000, the number of repetitions (N_I) to 4 and the hyperparameter α to 2. These values were chosen since they are typical for real datasets. The covariance matrix Σ is fixed and for each gene g the following steps were done.

1. c_g was sampled from an inverse gamma distribution according to the model specification.

2. A vector of $N_I = 4$ independent observations was drawn normal distribution with man value zero and variance one. This vector was then multiplied by the square-root matrix of Σ.

3. If this particular gene was selected to be regulated, then the absolute mean value for each of the N_I elements was drawn from a uniform distribution between 0 and 2.

5% of the genes were randomly selected and set to be upregulated. Analogously, 5% were downregulated resulting in totaly 10% regulated genes. It should be noted that it is only the total number of regulated genes had an impact on the performance for the different methods, not the number of upregulated genes compared to the number of downregulated genes.

Four cases, all with different covariance matrices, were studied. In the first case, all of the repetitions had variances equal to 1 and there were no correlations, thus Σ was an identity matrix. The ROC curves produced by the simulated data can be seen in the upper part of Figure 1. WAME and LIMMA performs best, closely followed by the penalized t-statistic. Note that WAME and LIMMA have almost identical performance in this case and, as mention above, this was expected since the weighted moderated t-statistic and the moderated t-statistic are almost equivalent for this setting. Another interesting detail is the low performance of the t-statistic due to its instability issues when only few repetitions are used.

In the second case, different variances were introduced. Σ was again a diagonal matrix but with the values 0.5, 1, 1.5 and 2 on the diagonal, thus all correlations were again zero. The ROC curves can be seen in the lower part of Figure 1. As before, WAME and LIMMA are the methods that performs best, but in this case, WAME performs better since it put less weight on the repetitions with high variance.

To investigate the impact of correlations, the third case used

$$\Sigma = \begin{pmatrix} 1.0 & 0.4 & 0.2 & 0.0 \\ 0.4 & 1.0 & 0.4 & 0.2 \\ 0.2 & 0.4 & 1.0 & 0.4 \\ 0.0 & 0.2 & 0.4 & 1.0 \end{pmatrix}. \tag{6}$$

This corresponds to a case when there are both high and low correlations between the repetitions. The upper part of Figure 2 shows that WAME performs slightly better than both LIMMA and the penalized t-statistic since it estimates the correlations and takes them into account.

Finally, in the fourth case both different variances and correlations were included. The variances and correlations were identical to the ones in the second and third cases respectively, i.e. variances of 0.5, 1.0, 1.5, 2.0 and correlations of 0, 0.2 and 0.4, the latter placed according to (6). The result can be seen in the lower part of Figure 2. Here, the largest advantage of using WAME can be seen. For a rejection threshold such that half of the selected genes are true positives, using WAME results in almost a third less false positives which can correspond to hundreds of genes.

All four simulations show that WAME and its weighted moderated t-statistic perform as least as well as the moderated and penalized t-statistics. In the case of both different variances and correlations between the repetitions, WAME performs clearly better than all of the included methods. Both the average fold-change and the ordinary t-statistic have poor performance in the current setting with only four repetitions.

5.2 Evaluation of the point estimation of Σ

The estimation of Σ is one of the crucial steps when applying WAME since errors made will affect estimates of other entities such as α and the weighted mean value $\bar{x}_g^w$. The resulting precision and accuracy when numerical maximum likelihood is applied to the distribution in equation (2) are therefore interesting questions, both when the model assumptions hold and when they are violated. In an attempt to partially answer these questions, Σ was estimated from different simulated datasets and the results were compared to the true values. The datasets were created according to the description in the previous section and the same parameters were used, i.e. $N_G = 10000$, $N_I = 4$ and $\alpha = 2$. In total, five different cases were examined, listed in Table 1. As in the previous section, 100 datasets were simulated for each setting and for each such dataset the covariance matrix Σ and the hyperparameter α were estimated according to Section 4. Table 2 summarises the

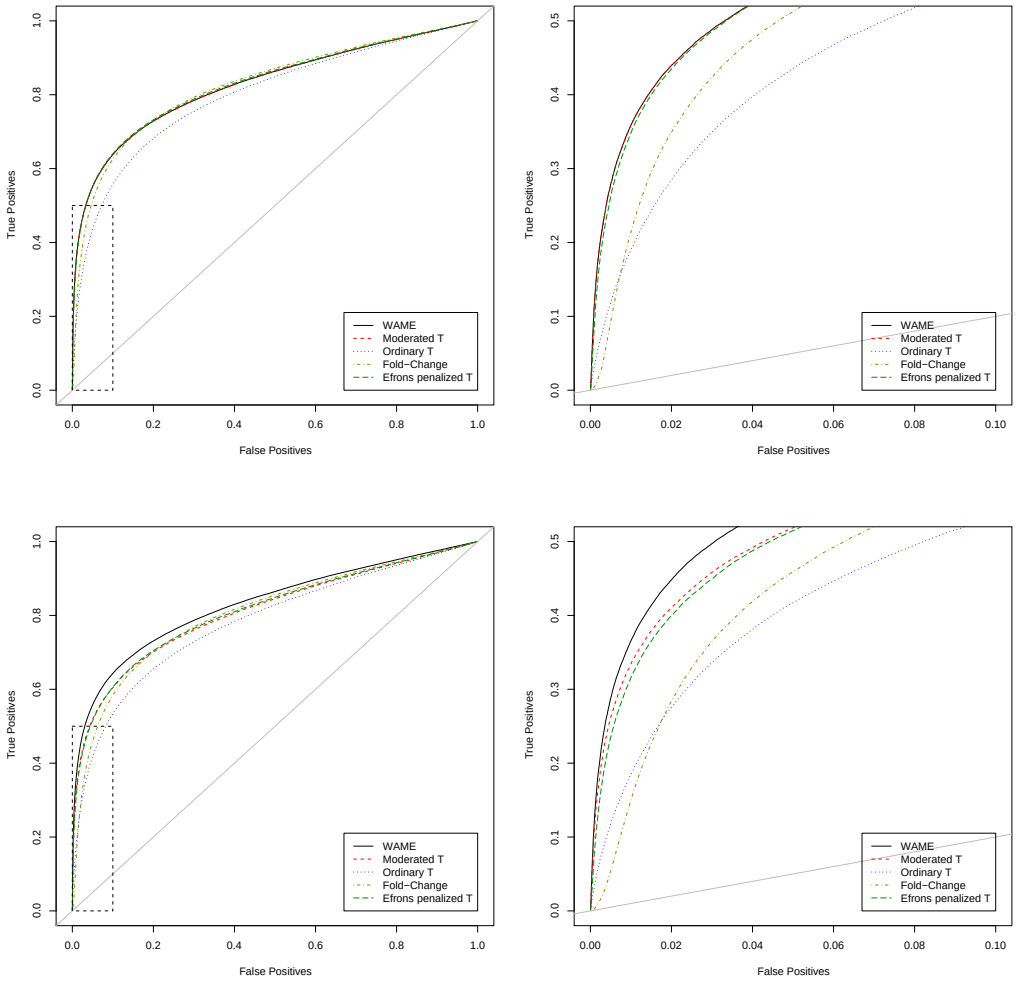


Figure 1: ROC curves from simulated data. The pair at the top, from the first case, show the performance of the evaluated methods on data with equal variances of 1 for all replicates and no correlations. The pair at the bottom, from the second case, analogously show the performance on data with different variances of 0.5, 1, 1.5, 2 and no correlations. The parameters used for these two simulations were as follows. $N_G = 10000$, $N_I = 4$, $\alpha = 2$ and 10% of the genes were regulated. The figures to the right are magnifications of the dashed boxes to the left.

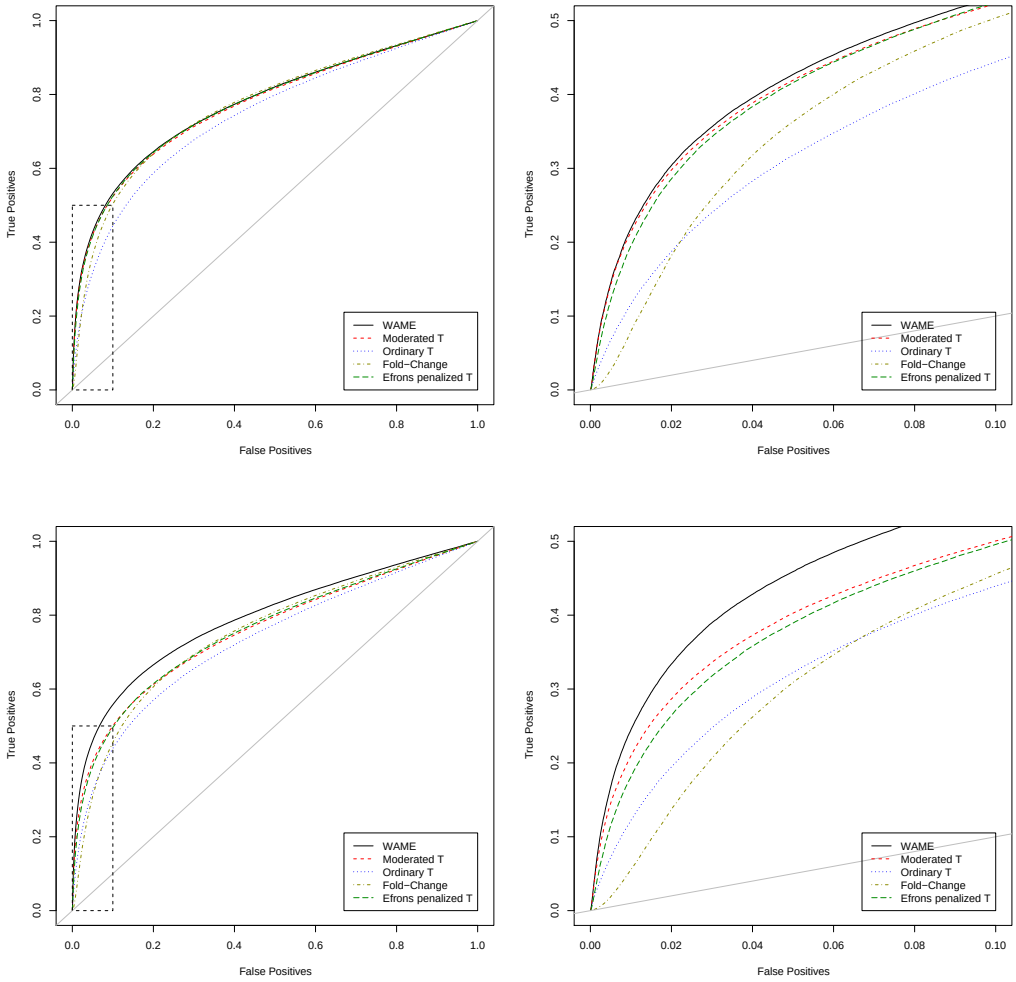


Figure 2: ROC curves from simulated data. The pair at the top, from the third case, show the performance of the evaluated methods on data with equal variances of 1 for all replicates and correlations of 0, 0.2 and 0.4, placed according to (6). The pair at the bottom, from the fourth case, analogously show the performance on data with different variances of 0.5, 1, 1.5, 2 and correlations of 0, 0.2 and 0.4, placed according to (6). The parameters used for these two simulations were as follows. $N_G = 10000$, $N_I = 4$, $\alpha = 2$ and 10% of the genes were regulated. The figures to the right are magnifications of the dashed boxes to the left.

Case	Correlation	Heavy tails	Regulated genes	Filter
I	No	No	None	No
II	Yes	No	None	No
III	Yes	Yes	None	No
IV	Yes	No	Yes, 10%	No
V	Yes	No	Yes, 10%	Yes, 5% removed.

Table 1: Descriptions of the five different settings used in this simulation study. When correlations are used, they follow the structure in equation (6).

result where the true value of Σ, the mean value of the estimated Σ as well as the standard deviations are listed. It should be noted that in all cases, except for case **III**, α is estimated with high accuracy and precision.

In the first two cases (**I** and **II**), the covariance matrix was estimated without any bias and with low standard deviation showing that the methods is accurate under the model assumptions. In case **III** the normal distribution was substituted against a t-distribution with 5 degrees of freedom, having substantially heavier tails. The estimated Σ seems to be slightly biased toward higher variances and α was estimated to 1.55 instead of 2. This pattern was also seen when the degrees of freedom were increased to 10 and 15 (results not shown). In case **IV** 10% of the genes were set to be regulated and since no differentially expressed genes are assumed, the regulation leads to positive correlations and increased variance estimates. Having 10% of the genes regulated is a rather high number, but not extreme. Therefore, a filter was applied to minimise the impact of regulated genes on the estimation of the covariance matrix. For each gene g, the filter calculates the minimal absolute value of the fold change, which will be denoted $X_{g,min}$. Removing the top 5% of the genes with highest $X_{g,min}$ gave a much better estimate of Σ, which is included as case **V**. Note that the genes were only removed for the estimation of Σ (not for the estimation of α and λ) and that the number 5% depends on several parameters, such as the total number of regulated genes and the covariance matrix itself. The results of the filtering procedure on real data is presented in the next section.

6 Results from real data

WAME was run on three real data sets: the ischemic part of the dataset of Hall et al. (2004), the dataset of Benson et al. (2004) (henceforth referred to

	True Σ				Mean estimated Σ				Standard deviation			
I	0.50	0.00	0.00	0.00	0.50	0.00	-0.00	-0.00	0.01	0.01	0.01	0.01
	0.00	1.00	0.00	0.00	*0.00*	1.01	-0.00	0.00	*0.01*	0.04	0.02	0.01
	0.00	*0.00*	1.50	0.00	*-0.00*	*-0.00*	1.51	-0.00	*0.02*	*0.02*	0.05	0.02
	0.00	*0.00*	*0.00*	2.00	*-0.00*	*0.00*	*-0.00*	2.02	*0.01*	*0.01*	*0.01*	0.07
II	0.50	0.28	0.17	0.00	0.50	0.28	0.17	0.00	0.02	0.01	0.01	0.01
	0.40	1.00	0.49	0.28	*0.40*	1.00	0.50	0.29	*0.01*	0.04	0.02	0.03
	0.20	*0.40*	1.50	0.69	*0.20*	*0.40*	1.51	0.70	*0.01*	*0.01*	0.06	0.04
	0.00	*0.20*	*0.40*	2.00	*0.00*	*0.20*	*0.40*	2.00	*0.01*	*0.01*	*0.01*	0.11
III	0.50	0.28	0.17	0.00	0.51	0.29	0.18	-0.00	0.02	0.01	0.01	0.01
	0.40	1.00	0.49	0.28	*0.40*	1.01	0.50	0.28	*0.01*	0.04	0.02	0.02
	0.20	*0.40*	1.50	0.69	*0.20*	*0.40*	1.52	0.70	*0.01*	*0.01*	0.05	0.03
	0.00	*0.20*	*0.40*	2.00	*-0.00*	*0.20*	*0.40*	2.03	*0.01*	*0.01*	*0.01*	0.07
IV	0.50	0.28	0.17	0.00	0.61	0.39	0.28	0.11	0.02	0.02	0.02	0.01
	0.40	1.00	0.49	0.28	*0.48*	1.11	0.60	0.39	*0.01*	0.04	0.03	0.01
	0.20	*0.40*	1.50	0.69	*0.28*	*0.45*	1.61	0.80	*0.01*	*0.01*	0.06	0.04
	0.00	*0.20*	*0.40*	2.00	*0.10*	*0.25*	*0.43*	2.11	*0.01*	*0.01*	*0.01*	0.08
V	0.50	0.28	0.17	0.00	0.46	0.21	0.11	-0.02	0.01	0.01	0.01	0.02
	0.40	1.00	0.49	0.28	*0.33*	0.90	0.38	0.22	*0.01*	0.02	0.02	0.02
	0.20	*0.40*	1.50	0.69	*0.14*	*0.34*	1.39	0.59	*0.01*	*0.02*	0.06	0.03
	0.00	*0.20*	*0.40*	2.00	*-0.02*	*0.16*	*0.36*	1.93	*0.02*	*0.01*	*0.01*	0.07

Table 2: Result from the estimations of Σ from each of the five different cases. Correlations are shown in italic and covariances in non-italic. The parameter values used were $N_G = 10000$, $N_I = 4$ and $\alpha = 2$. The mean values and sample standard deviations were calculated from the result of 100 simulated dataset. Refer to Table 1 for a description of the different cases.

as the *Cardiac* and *Polyp* datasets, respectively) and the *Swirl* dataset (described in chapter 3.3 of Dudoit and Yang, 2003). These datasets represent microarray experiments with different characteristics; different laboratories, both two-colour cDNA and one-channel oligonucleotide (Affymetrix) arrays, different tissues and two different species (human and zebrafish). The Cardiac and Swirl datasets are publicly available.

The Cardiac dataset is described to have been strictly quality controlled by a combination of several available methods. The dataset is therefore interesting to examine to see if WAME detects relevant differences in quality even in an example of a quality controlled, publicly available dataset. The Polyp dataset includes one biopsy that was previously thought to be an outlier and therefore discarded, thus providing a case with one seemingly lesser quality to be detected. In the Swirl dataset, two highly differentially expressed genes exist. Therefore, it is of interest to check that those genes are highly ranked by WAME. Furthermore, the Swirl dataset has been analysed in e.g. (Smyth, 2004).

6.1 Cardiac dataset

In the public dataset from Hall et al. (2004), heart biopsies from 19 patients with heart failure were harvested before and after mechanical support with a ventricular assist device. The aim of the study was to "define critical regulatory genes governing myocardial remodelling in response to significant reductions in wall stress", where a first step was to identify differentially expressed genes between the two conditions.

Affymetrix one-channel oligonucleotide arrays of type HG-U133A were used in the study, each containing 22283 probe-sets. The quality of the arrays was controlled using quality measures recommended by Affymetrix as well as by the program Gene Expressionist (GeneData, Basel, Switzerland). The quality of the different lab steps leading to the actual hybridisations were controlled using standard methods. The 19 patients were divided into three groups: ischemic (5 patients), acute myocardial infarction (6 patients) and non-ischemic (8 patients). The ischemic group was the smallest and consequently the one where quality variations might make the biggest difference. It was therefore chosen for further examination using WAME, to see if relevant quality variations could be detected despite the close quality monitoring.

The dataset was retrieved in raw .CEL-format from the public repository Gene Expression Omnibus (Edgar et al., 2002). The .CEL-files were subsequently processed using RMA (Irizarry et al., 2003) on all the arrays of the 19

patients simultaneously. Patient-wise $\log_2$-ratios of the five ischemic patients were then formed by taking pairwise differences of the $\log_2$ measurements before and after implant.

Applying WAME to the patient-wise $\log_2$-ratios provided interesting results. The estimated covariance matrix (see Table 3) suggests that two of the five patients (I13 and I7) were substantially more variable than the others, while the correlations between patients were rather limited. These numbers seem credible when examining Figure 3, where for each pair of patients, the respective $\log_2$-ratios of all genes were plotted against each other. The plots clearly imply that the observations of the two patients in question (I13 and I7) are more variable than the others.

The corresponding weights, derived from the estimated covariance matrix Σ, are shown in Table 4. As was discussed in Sections 4.1 and 5.2, when estimating Σ all genes are assumed to be non-differentially expressed. To examine the impact of potentially regulated genes on the estimation of Σ, the analysis was redone, removing genes with high lowest absolute $\log_2$-ratio in the estimation of Σ, as described in Section 5.2. The individual elements of the estimated covariance matrix and of α changed only slightly, even when as much as 50% of the data was removed (data not shown). This is reflected in the stable weights in Table 4.

	Patient				
Patient	I12	I13	I4	I7	I8
I12	0.046	0.003	0.001	0.012	0.002
I13	*0.033*	0.196	-0.014	0.007	-0.001
I4	*0.023*	*-0.126*	0.065	0.013	0.002
I7	*0.111*	*0.030*	*0.102*	0.258	-0.017
I8	*0.040*	*-0.011*	*0.038*	*-0.152*	0.047

Table 3: Estimated covariance-correlation matrix, Σ, for patients in the Cardiac dataset. (Correlations in italic, covariances in non-italic.)

The hyperparameter α related to the spread of the gene-wise variance components, c_g, was estimated to 1.92, giving a thick tail for the prior distribution. Thus removing c_g by transformation when estimating Σ (Section 4.1) is justified.

Inspecting the fitted distribution of S_g given $\alpha = 1.92$ against the empirical distribution of S_g reveals a good fit (see Figure 4), implying that the family of inverse gamma prior distributions is rich enough for this dataset.

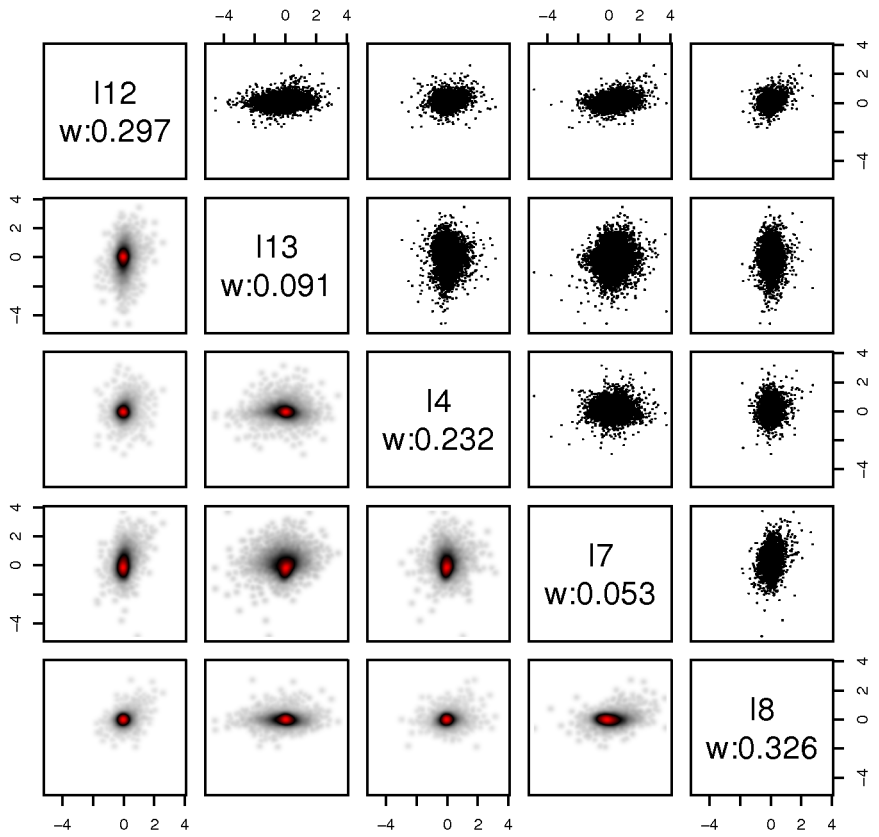


Figure 3: Pair-wise plots of the $\log_2$-ratios of the patients in the Cardiac dataset. The plots to the lower-left show two-dimensional kernel density estimates of the distribution of $\log_2$-ratios in each pair of patients. This provides information in the central areas where the corresponding scatterplots are solid black (cf. Figure 6 in Huber et al., 2003). The colour-scale is, in increasing level of density: white, grey, black and red.

	Patient				
Removed genes	I12	I13	I4	I7	I8
none	0.297	0.091	0.232	0.053	0.326
5%	0.301	0.089	0.233	0.054	0.323
10%	0.303	0.087	0.235	0.053	0.321
50%	0.323	0.082	0.240	0.046	0.308

Table 4: Weights for patients in the Cardiac dataset. Different numbers of potentially regulated genes were removed in the estimation of Σ, to check their influence. Potential regulation was measured by minimal $\log_2$-ratio among the patients.

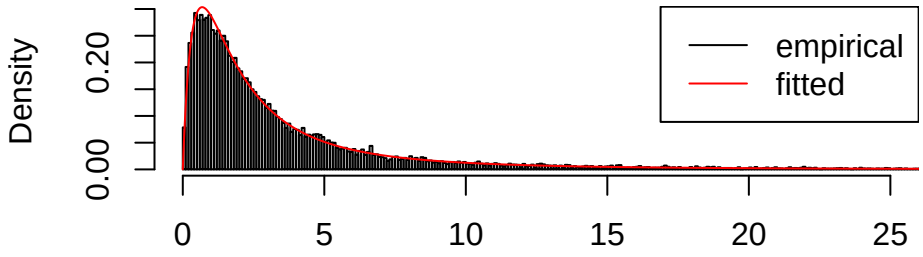


Figure 4: Empirical distribution of S_g in the Cardiac dataset, together with the density of S_g given $\alpha = 1.92$.

Examining the observed values of the statistic, T_g, compared to the expected null distribution reveals a good overall concordance (see Figure 5). Some genes have a larger t_g than can be explained by the null distribution, which points toward some of them being up-regulated by the treatment (see the qq-plot in Figure 5).

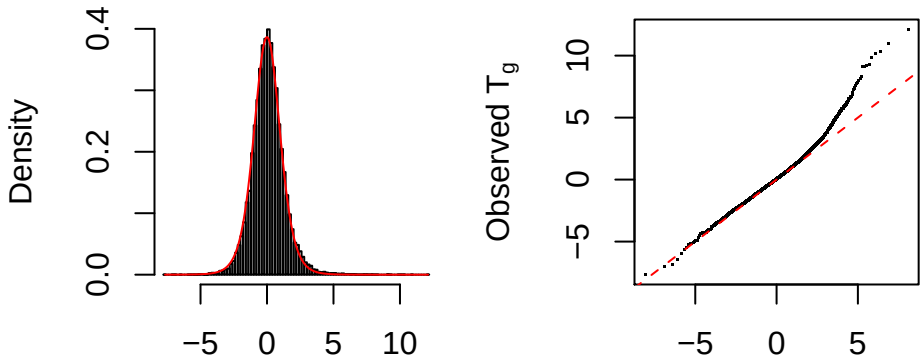


Figure 5: To the left, a histogram of the observed statistic, t_g, together with the density of the null distribution (in red), in the Cardiac dataset. To the right, a quantile-quantile plot where the observed values of T_g are paired with the quantiles of T_g under the null hypothesis. The points generally follow the identity line well, showing good concordance with the null distribution. For high positive T_g, the observed values exceed the predicted ones, pointing at the existence of up-regulated genes.

6.2 Polyp dataset

In the dataset from Benson et al. (2004), biopsies from nasal polyps of five patients were taken before and after treatment with local glucocorticoids. The goal was to examine closer the mechanisms behind the effect of the treatment and one step was to identify differentially expressed genes. Technical duplicates stemming from the same extracted RNA were run for each biopsy on Affymetrix HG-U133A arrays, forming a dataset of 20 arrays and 22283 probe-sets.

Looking at pair-wise scatterplots between arrays, the arrays from before treatment of patient 2 showed substantially larger variation, comparing it to all other arrays, than any other array in the dataset. The biopsy in question was found to be considerably smaller than the others, providing possible

explanations by e.g. non-representativeness in tissue distribution. The data from patient 2 was therefore excluded in Benson et al. (2004).

WAME would preferably identify the patient 2 observation as having larger variation and downweight it. The data was processed using RMA (Irizarry et al., 2003) and the $\log_2$-ratio for each patient was formed by taking differences between the averages over the technical duplicates, before and after treatment, combining 4 arrays for each patient into one set of $\log_2$-ratios. Making one scatter plot of the two sets of $\log_2$-ratios for each pair of patients (Figure 6) clearly indicates that patient 2 is more variable than patients 1,3 and 5. Interestingly, the measurements from patients 1 and 2 seem to be highly correlated and patient 4 seems to have high variability.

Estimating the covariance matrix, Σ, the correlation between patients 1 and 2 is estimated to 0.82 (see Table 5), which is high but not unbelievable when studying Figure 6. The variance of patient 2 is furthermore estimated to four times that of patient 1. Examining the resulting weights, patient 2 actually receives a weight of -2% (see Table 6). The negativeness is a result of it's variance being much higher than that of patient 1, together with them being highly correlated. If negative weights are not satisfying, patient 2 could be removed altogether and the analysis rerun, which is essentially what was performed in Benson et al. (2004). Beside the result of the very low weight for patient 2, the other patients receive distinctly different weights, which is interesting.

	Patient				
Patient	1	2	3	4	5
1	0.300	0.493	0.000	-0.012	-0.067
2	*0.822*	1.200	0.004	0.041	-0.157
3	*0.002*	*0.012*	0.091	-0.071	-0.055
4	*-0.038*	*0.067*	*-0.417*	0.319	0.102
5	*-0.291*	*-0.340*	*-0.434*	*0.430*	0.178

Table 5: Estimated covariance-correlation matrix, Σ, for patients in the Polyp dataset. (Correlations in italic, covariances in non-italic.)

The hyperparameter α, related to the spread of the gene-wise variance components, c_g, was estimated to 1.97, giving infinite variance for the distribution of c_g. The fit of S_g given $\alpha = 1.97$ was very good (see Figure 10 in the Appendix).

Similar to in the Cardiac dataset, the weights were steadily estimated when potentially regulated genes were removed in the estimation of the co-

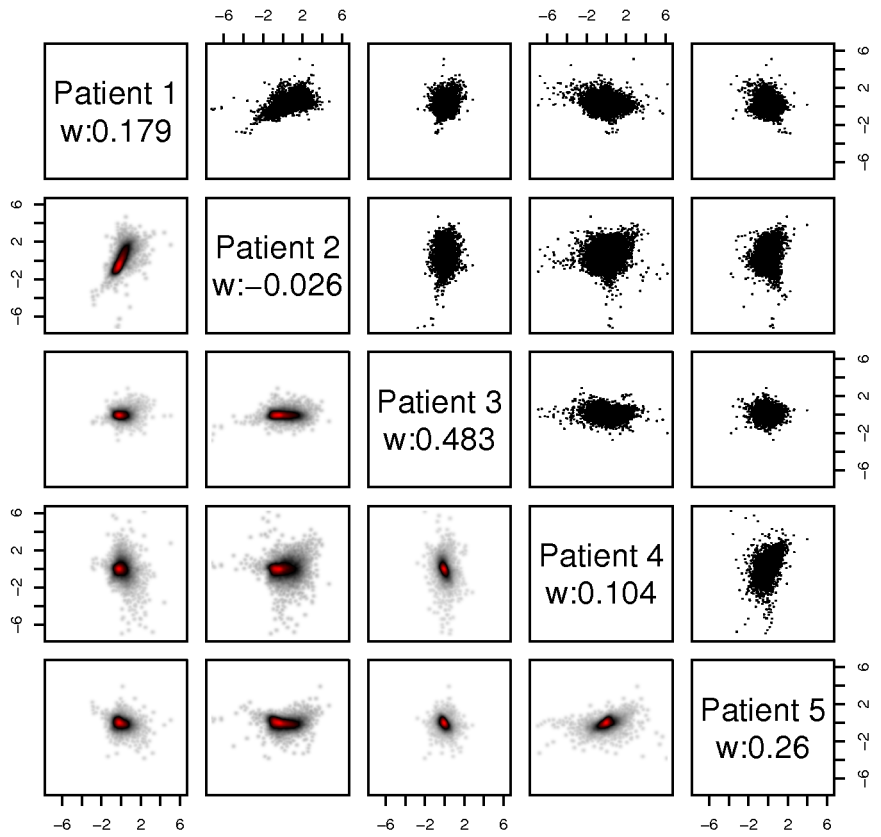


Figure 6: Pair-wise plots of the $\log_2$-ratios of the patients in the Polyp dataset. The plots to the lower-left show two-dimensional kernel density estimates of the distribution of $\log_2$-ratios in each pair of patients. This provides information in the central areas where the corresponding scatterplots are solid black (cf. Figure 6 in Huber et al., 2003). The colour-scale is, in increasing level of density: white, grey, black and red.

	Patient				
Removed genes	1	2	3	4	5
none	0.179	-0.026	0.483	0.104	0.260
5%	0.181	-0.025	0.481	0.104	0.259
10%	0.180	-0.024	0.482	0.103	0.259
50%	0.157	-0.015	0.506	0.100	0.252

Table 6: Weights for the patients in the Polyp dataset. Different numbers of potentially regulated genes were removed, to check their potential influence in the estimation of Σ. Potential regulation was measured by minimal $\log_2$-ratio among the patients.

variance matrix Σ (see Table 6). However, the estimated correlations between patients 3, 4 and 5 were decreased somewhat. Removing 5% of the genes decreased those correlations by 0.03-0.04 and removing 10% decreased them by 0.06-0.07. The high correlation between patient 1 and 2 was only slightly decreased (<0.03), even when 50% of the genes were removed.

Examining the observed values of the statistic, T_g, compared to the expected null distribution (see Figure 7) reveals a good overall concordance. Some genes have a more extreme T_g than can be explained by the null distribution, which points toward many of them being regulated by the treatment (see the qq-plot in Figure 7).

6.3 Swirl dataset

In the Swirl experiment (described in chapter 3.3 of Dudoit and Yang, 2003), the authors are interested in identifying genes that are differentially expressed in zebrafish carrying a point mutated SRB2 gene, compared to ordinary, wild-type zebrafish. SRB2 and one of it's known targets, Dlx3 are expected to be highly differentially expressed in this experiment, thus these genes should be highly ranked using WAME. The Swirl dataset has been examined in Smyth (2004).

The dataset consists of four two-colour cDNA microarrays with 8448 spots, whose data are publicly available. We used standard pre-processing to compensate for effects such as background and dye bias (Background correction *subtract* and within-array normalisation *printtip loess* were used in the LIMMA package (Smyth et al., 2003)). It should be noted that between-array normalisation (Yang et al., 2002) was not performed in contrast to the analysis in Smyth (2004).

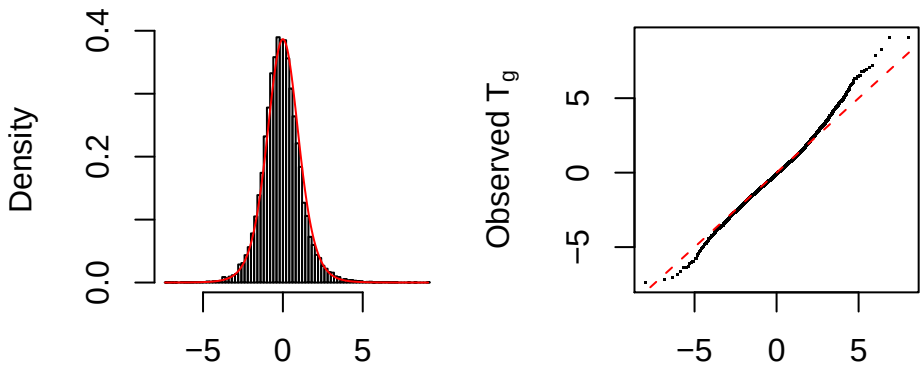


Figure 7: To the left, a histogram of the observed statistic, t_g, together with the density of the null distribution (in red), in the Polyp dataset. To the right, a quantile-quantile plot where the observed values of T_g are paired with the quantiles of T_g under the null hypothesis. The points generally follow the identity line well, showing good concordance with the null distribution. For extreme T_g, the observed values are more extreme than the predicted ones, pointing at the existence of regulated genes.

Making one scatter plot of the $\log_2$-ratios for each pair of arrays (Figure 8) indicates that array 2 is less variable than the others, while the genes with lowest $\log_2$-ratio on array 1 seem to be outliers, since they are not extreme in any other array. Examining the estimated covariance matrix (see Table 7), array 2 indeed receives the highest variance. In addition, there are substantial correlations between arrays 1-3, 2-4 and 3-4, which seems believable when examining the scatter-plots (Figure 8).

	Array			
Array	1	2	3	4
1	0.128	0.007	0.079	0.017
2	*0.066*	0.086	-0.002	0.038
3	*0.489*	*-0.017*	0.203	0.076
4	*0.136*	*0.371*	*0.482*	0.124

Table 7: Estimated covariance-correlation matrix, Σ, for the arrays in the Swirl dataset. (Correlations in italic, covariances in non-italic.)

When re-performing the estimation of Σ , removing potentially regulated genes (in analogy with the Polyp and Cardiac datasets), the correlations were

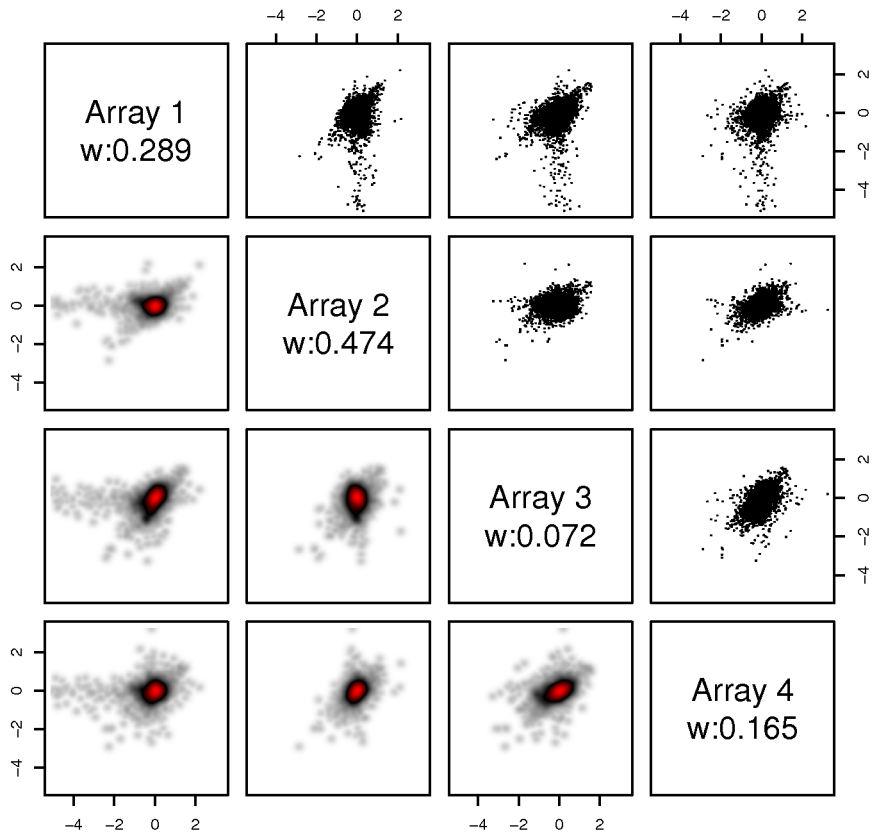


Figure 8: Pair-wise plots of the $\log_2$-ratios of the arrays in the Swirl dataset. The plots to the lower-left show two-dimensional kernel density estimates of the distribution of $\log_2$-ratios in each pair of patients. This provides information in the central areas where the corresponding scatterplots are solid black (cf. Figure 6 in Huber et al., 2003). The colour-scale is, in increasing level of density: white, grey, black and red.

decreased somewhat. Removing 5% of the genes decreased the three high correlations by 0.02-0.06, while removing 10% decreased them by 0.04-0.08. However, the corresponding weights only changed marginally (see Table 8).

	Array			
Removed genes	1	2	3	4
none	0.289	0.474	0.072	0.165
5%	0.288	0.469	0.076	0.166
10%	0.290	0.462	0.075	0.173
50%	0.282	0.447	0.087	0.184

Table 8: Weights for the arrays in the Swirl dataset. Different numbers of potentially regulated genes were removed, to check their potential influence in the estimation of Σ. Potential regulation was measured by minimal $\log_2$-ratio among the arrays.

The hyperparameter α was estimated to 1.89. Further analysing the dataset, the distribution of S_g fits the predicted distribution of S_g given $\alpha = 1.89$ well (see Figure 11 in in the Appendix). The observed values of the statistic, T_g, seem to fit the null distribution well (see Figure 9).

Since the point mutated gene, SRB2 and one of it's known targets, Dlx3, are expected to be highly differentially expressed, their actual ranking is of interest. In Table 9 below, the top 20 genes as ranked by WAME are listed. The values of some widely used statistics are included for comparison. The rankings by WAME and the moderated t-statistic (Smyth et al., 2003) are very similar, while the rankings by the ordinary t-statistic and the average $\log_2$-ratio (i.e. fold change) are rather different than the one by WAME, which was expected. All four spots for the two validated genes are included in WAME:s top 20 list (see Table 9).

7 Discussion

A drawback of the microarray technology is that it involves several consecutive steps, each exhibiting large quality variation. Thus there is a strong need for quality assessment and quality control to handle occurrences of poor quality. In this paper, we introduce a method called WAME for the analysis of paired microarray experiments, which aims at estimating array-wide quality deviations and integrates these quality estimates into the statistical analysis.

Name	ID	average $\log_2$-ratio	ordinary t-statistic	moderated t-statistic	WAME
fb85d05	18-F10	-2.66	-18.41	-20.79	-15.15
fb58g10	11-L19	-1.60	-14.32	-14.15	-11.51
control	Dlx3	-2.19	-15.91	-17.57	-11.17
control	Dlx3	-2.19	-13.58	-16.08	-9.84
fb24g06	3-D11	1.32	19.52	13.62	9.80
fb54e03	10-K5	-1.20	-25.74	-13.11	-9.66
fc22a09	27-E17	1.26	24.76	13.68	9.50
fb40h07	7-D14	1.35	14.15	12.69	9.12
fb85a01	18-E1	-1.29	-17.35	-13.01	-8.81
fb87f03	18-O6	-1.08	-27.90	-12.06	-8.80
fb37e11	6-G21	1.23	14.37	11.94	8.47
fb94h06	20-L12	1.28	15.41	12.54	8.46
fb87d12	18-N24	1.28	12.96	11.87	8.39
control	BMP2	-2.24	-8.63	-11.78	-8.33
fc10h09	24-H18	1.20	15.05	11.92	8.23
fb85f09	18-G18	1.29	11.50	11.38	8.22
control	BMP2	-2.33	-8.37	-11.58	-7.95
fb26b10	3-I20	1.09	15.50	11.17	7.81
fb37b09	6-E18	1.31	11.57	11.55	7.78
fc22f05	27-G10	-1.19	-10.42	-10.44	-7.70

Table 9: The top 20 most probably regulated genes in the Swirl dataset according to WAME.

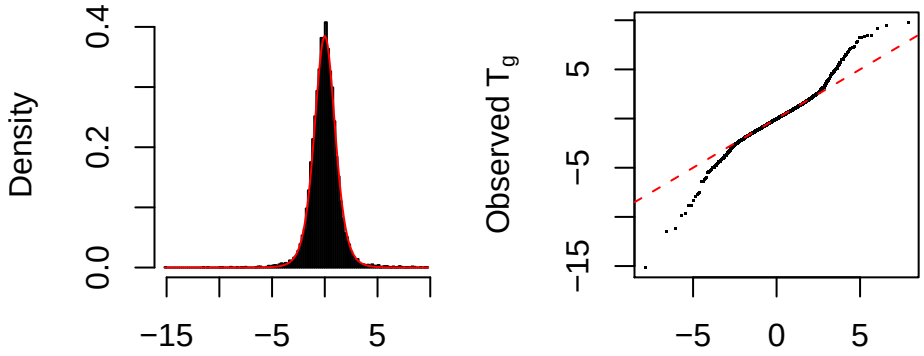


Figure 9: To the left, a histogram of the observed statistic, t_g, together with the density of the null distribution (in red), in the Swirl dataset. To the right, a quantile-quantile plot where the observed values of T_g are paired with the quantiles of T_g under the null hypothesis. The points generally follow the identity line well, showing good concordance with the null distribution. For extreme T_g, the observed values are more extreme than the predicted ones, pointing at the existence of regulated genes.

The quality deviations are modelled as different variances for different repetitions (e.g. arrays) as well as correlations between them, thus catching both unequal precision and systematic errors. This is contained in a covariance matrix Σ. Genes have different variability (both biological and technical), which is modelled by a gene-specific variance scaling factor c_g. Given this structure, the pair-wise measured $\log_2$-ratios for each gene are assumed to be normally distributed. It should be straight-forward to incorporate exclusion of outlying gene-wise observations (e.g. spots) into the model. Including quantitative measures of quality of such observations, e.g. by a hierarchical variance component model (cf. Bakewell and Wit (2005)), would be interesting as feature work.

Estimation of the covariance matrix is non-trivial due to the gene-wise scalings and unknown differential expressions μ_g. Here, an assumption is made that most genes are not differentially expressed ($\mu_g = 0$) and a transformation is performed to remove the gene-wise scalings. Then, a scaled version of Σ is estimated using numerical maximum likelihood, based on the derived resulting distribution. The assumption of no differential expression somewhat limits the experimental setups that can be analysed. However, this is not as consequential as it might sound, since it is made by most of the pro-

cedures that have become *de facto* standard in the (preceding) normalisation step.

Since most microarray experiments contain only a few repetitions, the estimate of the gene-specific variance scaling factor c_g is imprecise, which can easily lead to false conclusions if not accounted for. Here, an empirical Bayes approach is taken where an inverse gamma prior distribution is assumed, in effect moderating extreme estimates (Baldi and Long, 2001; Lönnstedt and Speed, 2002; Smyth, 2004). The hyperparameter α determining the spread of the prior distribution is estimated from the data, by numerical maximum likelihood together with the scale of the previously estimated scaled Σ.

To identify differentially expressed genes a likelihood-ratio test is derived, resulting in the *weighted moderated t-statistic*, which is a generalisation of the moderated t-statistic in Smyth (2004). Here, the estimated covariance matrix Σ is used both to produce weights for the different repetitions and when estimating the gene-specific variances. The weighted mean is the estimate of differential expression with minimal variance.

As discussed above, array-wide quality deviations in all steps leading to the observed $\log_2$-ratios are estimated and incorporated into the analysis. However, the current paper is restricted to paired, two sample settings. Interesting directions for future work would be to generalise the model to include more experimental setups, e.g. non-paired two sample cases (e.g. from one-channel arrays) and time series. There, challenges would be to model a more general variance structure and to estimate the corresponding covariance matrix Σ with less restricting assumptions.

A simulation study was done to compare the performance of WAME to four published methods. On data without correlations and with equal variances between repetitions, WAME performs as well as the moderated t-statistic which assumes this structure. When correlations and/or unequal variances were included, WAME performs better than all the other methods. In one case, using WAME results in almost a third less false positives which can correspond to hundreds of genes. Evaluating the point estimation of the covariance matrix Σ revealed good precision and accuracy when no regulated genes were present. Including 10% regulated genes resulted in a bias, which was partly handled by removing genes likely to be regulated. In both cases estimation of the hyperparameter α was nearly unbiased and accurate. The estimate of Σ was unbiased when heavy tails was introduced in contrast to the estimate of α which was estimated to 1.55 instead of 2.

Three real datasets were analysed: the ischemic part of the dataset of Hall et al. (2004)(publicly available), the dataset of Benson et al. (2004) and the *Swirl* dataset (described in chapter 3.3 of Dudoit and Yang, 2003)(pub-

licly available). In all cases, relevant correlations and differences in precision between replicates were found, even in first dataset which had been quality controlled by a combination of several available methods. The exact origin of the correlations is an interesting, open question. In the second dataset one previously identified outlier was practically removed by WAME. In the Swirl dataset, expected differentially expressed genes are ranked among the top 20. Relevant empirical distributions showed good fit to the theoretic distributions, pointing toward the family of prior distributions for c_g being flexible enough and the normal assumption being satisfying.

To summarise, WAME estimates and integrates array-wide quality deviations into the analysis of paired microarray experiments. An empirical Bayes approach is used to moderate the gene-specific variance scale estimates, resulting in a weighted moderated t-statistic with a derived distribution. The performance of WAME has been evaluated on both simulated and real microarray data, with interesting results.

Acknowledgements

We would like to thank Mikael Benson, Lars Olaf Cardell, Lena Carlsson and Margareta Jernås for valuable discussions and access to the data from (Benson et al., 2004).

Appendix

Additional Figures

Mathematical details

We observe $\mathbf{X}_g = (X_{g1}, \ldots, X_{gN_I})$ where $g = 1, \ldots, N_G$. Let Σ be a covariance structure matrix for the N_I repetitions, c_g a set of gene-wise variance scaling factors and α a hyperparameter determining the shape of the prior distribution for c_g. Then for fixed μ_g, Σ and α,

$$\begin{aligned} c_g &\sim \Gamma^{-1}(\alpha, 1), \text{ and} \\ \mathbf{X}_g \mid c_g &\sim \mathbf{N}_{N_I}\left(\mu_g \mathbf{1}, c_g \Sigma\right) \end{aligned}$$

and all variables corresponding to different genes are assumed independent.

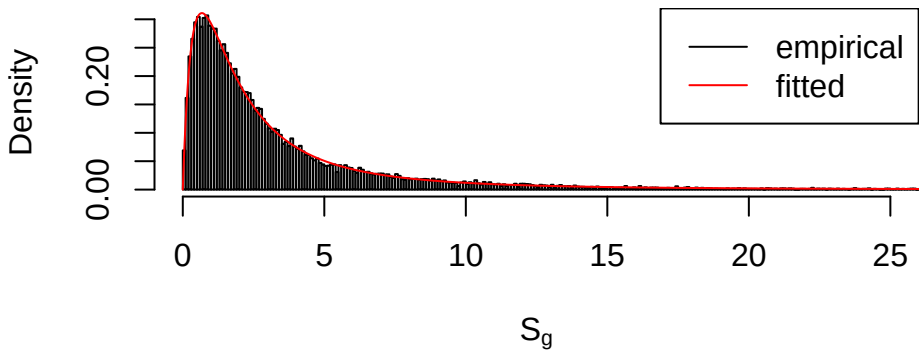


Figure 10: Empirical distribution of S_g in the Polyp dataset, together with the density of S_g given $\alpha = 1.97$.

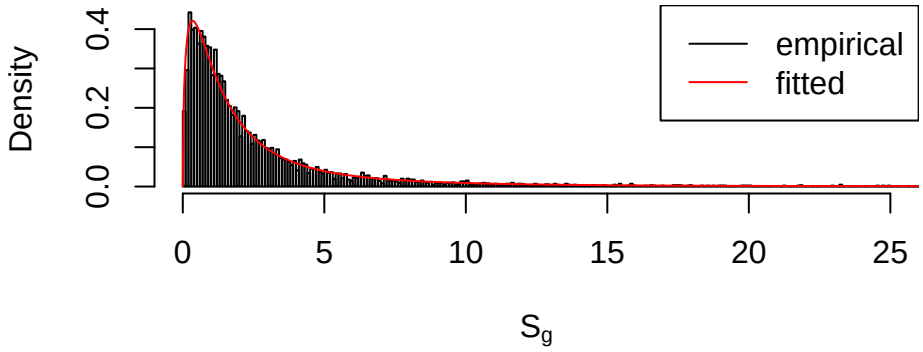


Figure 11: Empirical distribution of S_g in the Swirl dataset, together with the density of S_g given $\alpha = 1.89$.

Estimation of a scaled version of the matrix Σ

Assume that $\mu_g = 0$ for all g. Under this assumption, it is possible to derive a scale independent estimate of the covariance matrix Σ by a transformation of the vector $\mathbf{X}_g$. This is done as follows (the index g is dropped to increase the readability) Let $\mathbf{U} = (U_1, \ldots, U_{N_I})$ where

$$U_i = \begin{cases} X_1 & \text{if } i = 1 \\ X_i/X_1 & \text{if } 2 \leq i \leq N_I. \end{cases}$$

The inverse becomes

$$X_i = \begin{cases} U_1 & \text{if } i = 1 \\ U_i U_1 & \text{if } 2 \leq i \leq N_I. \end{cases}$$

and the Jacobian can be derived to

$$J(u_1, \ldots, u_{N_I}) = u_1^{N_I - 1},$$

so for $\mathbf{U} \in \mathbb{R}^{N_I}$ the density becomes

$$\begin{aligned} f_{\mathbf{U} \mid c,\Sigma}(\mathbf{u}) &= f_{\mathbf{X} \mid c,\Sigma}\left(\mathbf{x}(\mathbf{u})\right) |J(\mathbf{u})| \\ &= (2\pi)^{-N_I/2} \, c^{-N_I/2} \, |\Sigma|^{-1/2} \, |u_1|^{N_I - 1} e^{-\frac{u_1^2}{2c} v^{\mathrm{T}} \Sigma^{-1} v}. \end{aligned}$$

where $v = (1, u_2, \ldots, u_{N_i})^{\mathrm{T}}$. Integration over u_1 yields

$$\begin{aligned} f_{U_2,\ldots,U_{N_I} \mid \Sigma}(u_2, \ldots, u_{N_I} \mid \Sigma) &= \int_{-\infty}^{\infty} f_{\mathbf{U} \mid c,\Sigma}(\mathbf{u} \mid c, \Sigma) \, du_1 \\ &= C \, |\Sigma|^{-1/2} \left[v^{\mathrm{T}} \Sigma^{-1} v \right]^{-N_I/2}. \end{aligned} \tag{7}$$

C is a normalisation constant and v is defined as above. Observe that this density is scale invariant with respect to the parameter Σ in the sense that for any scalar λ,

$$f_{U_2,\ldots,U_{N_I} \mid \Sigma}(u_2, \ldots, u_{N_I} | \lambda\Sigma) = f_{U_2,\ldots,U_{N_I} \mid \Sigma}(u_2, \ldots, u_{N_I} | \Sigma).$$

Thus, it is also independent of c (as can be seen by looking at the function of the density) and under the assumption of independent genes, the log-likelihood function becomes

$$l(\Sigma | \mathbf{x}) = C' - \frac{N_G}{2} \log\left(|\Sigma|\right) - \frac{N_I}{2} \sum_{g=1}^{N_g} \log\left(v^{\mathrm{T}} \Sigma v\right),$$

where C' is a constant that is independent of Σ. Numerical maximisation yields a scaled version of Σ, denoted Σ^*.

Estimation of the hyperparameter α and the scale λ

From the model assumptions, we know that

$$c_g \mid \alpha \sim \Gamma^{-1}(\alpha, 1).$$

Assume that Σ is known and define

$$S_g = (A\mathbf{X}_g)(A\Sigma A^{\mathrm{T}})^{-1}A\mathbf{X}_g,$$

where A is a contrast matrix, e.g., a matrix of dimension $N_I - 1 \times N_I$, with full rank and with each row sum equal to 0. It follows that

$$S_g \sim c_g \times \chi^2_{N_I-1}.$$

The unconditional distribution of S_g can be derived by integrating over c_g, i.e.,

$$\begin{aligned} f_{S_g \mid \alpha}(s_g) &= \int_0^\infty f_{S_g \mid c_g}(s) f_{c_g \mid \alpha}(c_g)\, dc_g \\ &= \frac{1}{2}\frac{(s/2)^{(N_I-1)/2-1}}{\Gamma(\alpha)\,\Gamma((N_I-1)/2)} \int_0^\infty c^{-\alpha-(N_I-1)/2-1} e^{-(s/2+1)}\, dc_g \\ &= \frac{1}{2}\frac{\Gamma(\alpha+(N_I-1)/2)}{\Gamma(\alpha)\Gamma((N_I-1)/2)} \frac{(s/2)^{(N_I-1)/2-1}}{[1+s/2]^{\alpha+(N_I-1)/2}}. \end{aligned}$$

This is a beta prime distribution (also called a beta distribution of the second kind) (Johnson et al., 1995) with parameters $N_I - 1$ and α which is denoted by $\beta'(N_I - 1, \alpha)$. Since only a scaled version of Σ, denoted Σ^*, is assumed known from the primary estimation step, following entities are defined. Let

$$\begin{aligned} \Sigma^* &= \lambda\Sigma \\ S_g^* &= (A\mathbf{X}_g)^{\mathrm{T}}(A\Sigma^* A^{\mathrm{T}})^{-1}A\mathbf{X}_G = S_g/\lambda, \end{aligned}$$

where λ in the unknown scale for Σ^*. It follows that

$$S_g^* \sim 2/\lambda \times \beta'(N_I - 1, \alpha).$$

The likelihood function can be simplified to

$$\begin{aligned} l(\alpha, \lambda | \{s_g\}_{g=1}^{N_G}) = C + N_G\,[&(N_I-1)/2\log(\lambda) + \log\Gamma(\alpha+(N_I-1)/2) - \log\Gamma(\alpha)] \\ &- (\alpha+(N_I-1)/2)\sum_{g=1}^{N_G}\log(s_g\lambda/2+1). \end{aligned}$$

Numerical maximum likelihood is used to estimate α and λ, which together with Σ^* can be used to calculate an estimate for Σ.

Inference about μ_g

The hypotheses that are interesting to test are if different genes are regulated or not, that is for each g,

$$\begin{aligned} H_0 &: \text{ gene } g \text{ is not regulated } (\mu_g = 0) \\ H_A &: \text{ gene } g \text{ is regulated } (\mu_g \neq 0). \end{aligned}$$

To test these hypotheses a maximum likelihood ratio (LRT) test is derived. For each g, we reject H_0 if

$$\lambda(\mathbf{x}_g) = \frac{\sup_{\mu_g \neq 0} L(\mu_g|\mathbf{x}_g)}{L(0|\mathbf{x}_g)} \geq k,$$

where $1 \leq k < \infty$. The likelihood L can be calculated by integration over c_g, e.g.

$$\begin{aligned} L(\mu_g|\mathbf{x}) &= \int f_{\mathbf{x} \mid \mu_g, c_g}(\mathbf{x}) f_{c_g \mid \alpha}(c_g)\, dc_g \\ &= (2\pi)^{-N_I/2} |\Sigma|^{-1/2} \frac{\Gamma(N_I/2+\alpha)}{\Gamma(\alpha)} \left[\frac{(\mathbf{x}_g - \mu_g \mathbf{1})^{\mathrm{T}} \Sigma^{-1} (\mathbf{x}_g - \mu_g \mathbf{1})}{2} + 1 \right]^{-N_I/2-\alpha}. \end{aligned}$$

To calculate the numerator in the likelihood ratio we need to maximise L over μ_g, which is the same as minimising

$$(\mathbf{x}_g - \mu_g \mathbf{1})^{\mathrm{T}} \Sigma^{-1} (\mathbf{x}_g - \mu_g \mathbf{1})$$

A little algebra shows that this maximum corresponds to the argument

$$\hat{\mu}_g = \frac{\mathbf{1}^{\mathrm{T}} \Sigma^{-1}}{\mathbf{1}^{\mathrm{T}} \Sigma^{-1} \mathbf{1}} \mathbf{x}_g \, .$$

Observe that $\hat{\mu}_g$ is the weighted mean value with minimal variance for the covariance matrix Σ. We will use $\bar{x}_g^w$ to denote this weighted average. The maximum value of the likelihood function becomes

$$L(\bar{x}_g^w|\mathbf{x}_g) = (2\pi)^{-N_I/2} |\Sigma|^{-1/2} \frac{\Gamma(N_I/2+\alpha)}{\Gamma(\alpha)} \left[\frac{\mathbf{x}_g^{\mathrm{T}} \Sigma^{-1} \mathbf{x}_g - (\bar{x}_g^w)^2 \mathbf{1}^{\mathrm{T}} \Sigma^{-1} \mathbf{1}}{2} + 1 \right].$$

Using this, the likelihood ratio test statistic can be rewritten as

$$\begin{aligned}
\frac{L\left(\bar{x}_g^w|\mathbf{x}_g\right)}{L\left(0|\mathbf{x}_g\right)} &= \left[\frac{\mathbf{x}_g^{\mathrm{T}}\Sigma^{-1}\mathbf{x}_g + 2}{\mathbf{x}_g^{\mathrm{T}}\Sigma^{-1}\mathbf{x}_g - (\bar{x}_g^w)^2\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1} + 2}\right]^{N_I/2+\alpha} \\
&= \left[1 + \frac{(\bar{x}_g^w)^2\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}}{\mathbf{x}_g\Sigma^{-1}\mathbf{x}_g - (\bar{x}_g^w)^2\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1} + 2}\right]^{N_I/2+\alpha} \\
&= \left[1 + \frac{(\bar{x}_g^w)^2\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}}{(\mathbf{x}_g - (\bar{x}_g^w)\mathbf{1})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_g - (\bar{x}_g^w)\mathbf{1}) + 2}\right]^{N_I/2+\alpha} \\
&= \left[1 + \frac{(\bar{x}_g^w)^2\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}}{(A\mathbf{X}_g)^{\mathrm{T}}(A\Sigma A^{\mathrm{T}})^{-1}(A\mathbf{X}_g) + 2}\right]^{N_I+\alpha} \\
&= \left[1 + \frac{(\bar{x}_g^w)^2\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}}{s_g + 2}\right]^{N_I/2+\alpha} \geq k,
\end{aligned}$$

which is equivalent to

$$\frac{|\bar{x}_g^w|}{\sqrt{s_g + 2}} \geq k', \tag{8}$$

where $0 \leq k' < \infty$ is a new constant. A is the expressions above is the contrast matrix

$$A = \begin{pmatrix} 1 - w_1 & -w_2 & -w_3 & \dots & -w_{N_I} \\ -w_1 & 1 - w_2 & -w_3 & \dots & -w_{N_I} \\ \dots & \dots & \dots & \dots & \dots \\ -w_1 & -w_2 & -w_3 & \dots & 1 - w_{N_I} \end{pmatrix}$$

and w_i is the i:th element of the vector

$$\frac{\mathbf{1}^{\mathrm{T}}\Sigma^{-1}}{\mathbf{1}^{\mathrm{T}}\Sigma^{-1}\mathbf{1}}.$$

To derive the distribution of the statistic that corresponds to (8), we proceed as follows. Let

$$T_g = \frac{\bar{X}_g^w}{\sqrt{S_g + 2}}.$$

Then since

$$T_g = \frac{\bar{X}_g^w/\sqrt{c_g}}{\sqrt{S_g/c_g + 2/c_g}},$$

the nominator in this expression is independent of S_g and has the same normal distribution conditionally on all c_g (and thus also unconditionally

has this normal distribution), showing that the denominator in this ratio expression is independent of the nominator. Since furthermore a similar argument shows that also S_g/c_g and $2/c_g$ are independent, and since $2/c_g$ is χ^2 distributed with 2α degrees of freedom and S_g/c_g is χ^2 distributed with $N_I - 1$ degrees of freedom, the sum is distributed as a χ^2 with $N_I - 1 + 2\alpha$ degrees of freedom. Hence, it follows that T_g is a scaled and translated t-distribution with $N_I - 1 + 2\alpha$ degrees of freedom, namely,

$$T_g \mid \mu_g, \Sigma, \alpha \sim \frac{1}{\sqrt{N_I - 1 + 2\alpha}} \times t_{N_I - 1 + 2\alpha}(\mu_g).$$

We call T_g the weighted moderated t-statistic.

References

D.J. Bakewell and E. Wit. Weighted analysis of microarray gene expression using maximum-likelihood. *Bioinformatics*, 21:723–729, 2005.

P. Baldi and A.D. Long. A bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics*, 17(6):509–519, 2001.

M. Benson, L. Carlsson, M. Adner, M. Jernås, M. Rudemo, A. Sjögren, P.A. Svensson, R. Uddman, and Cardell L.O. Gene profiling reveals increased expression of uteroglobin and other anti-inflammatory genes in glucocorticoid-treated nasal polyps. *Journal of Allergy and Clinical Immunology*, 113(6):1011–1284, 2004.

D.-T. Chen. A graphical approach for quality control of oligonucleotide array data. *Journal of Biopharmaceutical Statistics*, 14(3):591–606, 2004.

S. Dudoit and J.Y.H. Yang. Bioconductor R packages for exploratory data analysis and normalization of cDNA microarray data. In G. Parmigiani, E.S. Garett, R.A. Irizarry, and S.L. Zeger, editors, *The Analysis of Gene Expression Data*. Springer, 2003.

C.I. Dumur, S. Nasim, A.M. Best, K.J. Archer, A.C. Ladd, V.R. Mas, D.S. Wilkinson, C.T. Garret, and A. Ferreira-Gonzalez. Evaluation of quality-control criteria for microarray gene expression analysis. *Clinical Chemistry*, 50(11):1994–2002, 2004.

R. Edgar, M. Domrachev, and A.E. Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1):207–210, 2002.

B. Efron, R. Tibshirani, J.D. Storey, and V. Tusher. Empirical Bayes Analysis of a Microarray Experiment. *Journal of the American Statistical Association*, 96(456):1151–1160, 2001.

J.L. Hall, S. Grindle, X. Han, D. Fermin, S. Park, Y. Chen, R.J. Bache, A. Mariash, Z. Guan, S. Ormaza, J. Thompson, J. Graziano, S.E. de Sam Lazaro, S. Pan, R.D. Simari, and L.W. Miller. Genomic profiling of the human heart before and after mechanical support with a ventricular assist device reveals alterations in vascular signaling networks. *Journal of Physiological Genomics*, 17(3):283–291, 2004. URL `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GDS558`.

S. Hautaniemi, H. Edgren, P. Vesanen, M. Wolf, A.-K. Jarvinen, O. Yli-Harja, J. Astola, K. Olli, and O. Monni. A novel strategy for microarray quality control using bayesian networks. *Bioinformatics*, 19(16):2031–2038, 2003.

W. Huber, A. von Heydebreck, and M. Vingron. Analysis of microarray gene expression data. In M. et al. Bishop, editor, *Handbook of Statistical Genetics, 2nd Edition.* John Wiley & Sons, 2003.

R.A. Irizarry, B.M. Bolstad, F. Collin, L.M. Cope, B. Hobbs, and T.P. Speed. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*, 31(4):e15, 2003.

K. Johnson and S. Lin. Qa/qc as a pressing need for microarray analysis: meeting report from camda'02. *Biotechniques*, 34(suppl):S62–S63, 3 2003.

N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions Volume 2*. Wiley, 1995.

I. Lönnstedt and T. Speed. Replicated microarray data. *Statistica Sinica*, 12 (1):31–46, 2002.

T. Park, S.-G. Yi, S. Lee, and J.K. Lee. Diagnostic plots for detecting outlying slides in a cdna microarray experiment. *BioTechniques*, 38(3): 463–471, 2005.

R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2004. URL `http://www.R-project.org`. ISBN 3-900051-00-3.

C.P. Robert. *The Bayesian Choice*. Springer, 2003.

M.M. Ryan, S.J. Huffaker, M.J. Webster, M. Wayland, T. Freeman, and S. Bahn. Application and optimization of microarray technologies for human postmortem brain studies. *Nucleic Acids Research*, 55(4):329–336, 2004.

L. Shi, W. Tong, F. Goodsaid, F.W. Frueh, H. Fang, T. Han, J.C. Fuscoe, and D.A. Casciano. QA/QC: Challenges and pitfalls facing the microarray community and regulatory agencies. *Expert Review of Molecular Diagnostics*, 4(6):761–777, 2004.

G. Smyth. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1), 2004.

G.K. Smyth, N.P. Thorne, and J. Wettenhall. *Limma: Linear Models for Microarray Data User's Guide. Software Manual available from http://www.bioconductor.org.*, 2003. URL `http://www.bioconductor.org`.

H. Tomita, M.P. Vawter, D.M. Walsh, S.J. Evans, P.V. Choudary, J. Li, K.M. Overman, M.E. Atz, R.M. Myers, E.G. Jones, S.J. Watson, H. Akil, and W.E. Bunney Jr. Effect of agonal and postmortem factors on gene expression profile: Quality control in microarray analyses of postmortem human brain. *Biological Psychiatry*, 55(4):346–352, 2004.

W. Tong, S. Harris, X. Cao, H. Fang, L. Shi, H. Sun, J. Fuscoe, A. Harris, H. Hong, Q. Xie, R. Perkins, and Casciano D. Development of public toxicogenomics software for microarray data management and analysis. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 549(1-2):241–253, 2004.

X. Wang, S. Ghosh, and S.W. Guo. Quantitative quality control in microarray image processing and data acquisition. *Nucleic acids research*, 29(15): e75, 2001.

X. Wang, M.J. Hessner, Y. Wu, N. Pati, and S. Ghosh. Quantitatative quality control in microarray experiments and the application in data filtering,

normalization and false positive rate prediction. *Bioinformatics*, 19(11): 1341–1347, 2003.

Y.H. Yang, S. Dudoit, P. Luu, D. Lin, V. Peng, J. Ngai, and T.P. Speed. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, 30(4):e15, 2002.