Chemometrics and Intelligent Laboratory Systems 98 (2009) 38–44

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Robust ANOVA for microarray data

J.R. de Haan [a], S. Bauerschmidt [b], R.C. van Schaik [b,d], E. Piek [c], L.M.C. Buydens [a], R. Wehrens [a,*]

[a] Institute for Molecules and Materials, Analytical Chemistry, Radboud University Nijmegen, Toernooiveld 1, 6525 ED, Nijmegen, The Netherlands
[b] Schering-Plough, Molenstraat 110, 5340 BH, Oss, The Netherlands
[c] Department of Applied Biology, Faculty of Science, Radboud University Nijmegen, Toernooiveld 1, 6525 ED, Nijmegen, The Netherlands
[d] Centre for Molecular and Biomolecular Informatics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen, Geert Grooteplein 28, 6525 GA, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 30 January 2009
Received in revised form 16 April 2009
Accepted 30 April 2009
Available online 15 May 2009

Keywords:
ANOVA
Robust analysis
PCA
ROC curves
Microarray

ABSTRACT

Analysis of Variance (ANOVA) separates the effects of different factors in a dataset. Typical examples for gene microarray data are the factors time and treatment. This separation can improve the interpretability of the results. However, the main effects and interactions, calculated in ANOVA, can be heavily influenced by outliers, large numbers of non-expressed genes with noise, and the heavy-tailedness of the distribution of expression values. Robust methods are less affected by these and will improve the analysis.
In this paper, several methods to perform robust nonparametric ANOVA are applied to a large multi-treatment time series dataset. The results are compared with the results obtained with parametric ANOVA using Procrustes analysis. A further comparison is made by Gene Ontology (GO) enrichment analysis of groups of genes identified as significant by inspection of the interaction terms in ANOVA. It is shown that there are significant differences in the estimates of main effects and gene–treatment interactions. ROC curves show an improved representation of current biological knowledge for one particular robust form of ANOVA, using a combination of rank transformed data, with the median as location parameter.

1. Introduction

The biological interpretation of high throughput gene expression experiments is difficult, because interesting information can be obscured by the effect of experimental factors on the data. A popular statistical method in the analysis of microarray data is Analysis of Variance (ANOVA). By applying ANOVA, a model can be made which separates the experimental, biological and residual effects. ANOVA has been applied to analyse microarray data in several research fields, for instance to investigate diseases, nutrition, microbiology, plant biology and many others [1–3]. The use of ANOVA for microarrays has been proposed by Kerr et al. [4] with a fixed effect model. In this model the array effect, dye effect, treatment effect, gene effect and interaction effects are modeled for two colour microarrays. The model was split in two stages by Lee and Whitmore [5]. The objective of the first stage is to normalize and filter out dye and array effects with ANOVA. The second stage is a per gene ANOVA on the remaining effects. Wolfinger et al. [6] assumed the array effect to be random in a model similar to the approach of Lee et al. This results in a two-stage mixed-effect model.

Regular parametric ANOVA assumes normally distributed data. This condition is rarely met for microarray data; Durbin et al. [7] have shown that in many cases a mixture of a normal and a lognormal distribution is more suited:

$$\gamma = \alpha + \mu e^{\eta} + \epsilon \tag{1}$$

Here, α is the background signal, μ is the real expression, and η and ϵ are normally distributed error terms. One could imagine that genes that do not respond to a certain treatment lead to a normal component in the overall distribution, whereas genes that do respond provide the lognormal component. Moreover, there is always the possibility of outliers because of, e.g., measurement errors.

All this implies that the estimates for the main effects and the interactions are influenced by observations that are outlying, either because of true biological differences in gene behaviour, or because of measurement errors. Durbin et al. [7] stated that before further statistical analysis can take place microarray data should be transformed to approach normality, and several transformations have been proposed in literature [8–10].

Here, we take a different approach, and concentrate on nonparametric forms of ANOVA that do not rely on distributional assumptions of the data. Nonparametric methods have been applied to microarray data before, for instance to identify differentially expressed genes [11–13], with good results. We compare three different methods for robust ANOVA in the analysis of microarray data. The first method simply replaces the non-robust mean with the robust median and proceeds as in an ordinary ANOVA. Two further methods are based on rank transformations.

* Corresponding author.
E-mail address: R.Wehrens@science.ru.nl (R. Wehrens).

0169-7439/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2009.04.011

The three methods are compared using a large time series dataset with several different treatments, designed to investigate the differentiation of mesenchymal stem cells to osteoblasts in the process of osteogenesis. Rather than focussing on the significance of individual terms, we treat the ANOVA as a means to concentrate on the interesting information. For this paper, we take the interaction between gene and treatment because it is biologically the most interesting term, and use Principal Component Analysis (PCA) for the visualisation of this interaction matrix, as in [14]. In the field of agricultural sciences this method is known as AMMI (Additive Main effect Multiplicative Interaction) [15,16]. The ANOVA–PCA combination can be used to select interesting genes based on the scores of the gene–treatment interaction [14].

Procrustes analysis [17] is used to compare the results from the original approach with the robust approaches. Furthermore, results from Gene Ontology (GO) [18] enrichment calculations [19] are used to draw Receiver Operating Characteristics (ROC) curves. These ROC curves can be used to assess the agreement between current knowledge about a biological system and the enrichment results from the selected genes [20]. Biological knowledge about the role of genes known in stem cell differentiation is used and discussed in the Supplementary material.

2. Methods and analysis

2.1. The mesenchymal stem cell dataset

The microarray experiment used in the examples presented here was performed on human Mesenchymal Stem Cells (hMSCs), triggered to undergo osteogenic differentiation (E. Piek et al., manuscript in preparation, [14]). Dexamethasone induces the differentiation of hMSCs cells to osteoblasts. It is known that BMP2 and Vitamin D3 can potentiate osteogenesis in combination with dexamethasone [21]. A time series experiment was performed and the expression measurements were taken at 10 time points (at 1, 3, 6, 12, 24, 48, 72, 120, 192 and 288 h after onset of treatment). As onset of osteogenic differentiation is expected during the first 24 h of treatment, frequent sampling was performed within this time frame. In addition to sampling multiple time points during the differentiation process, another factor of interest is introduced to the experiment by inducing osteogenesis of the hMSCs with different treatments. The three treatments are named after the substances added to the culture medium. For the first treatment (VIT) a combination of Vitamin D3 and dexamethasone is used. The second treatment (BMP) consists of a combination of Bone Morphogenetic Protein 2 and dexamethasone. The third treatment (DEX) consists of dexamethasone only. Furthermore an untreated sample (UNT) is measured as a control at each time point. Thus, four time profiles are available for each gene on the microarrays. The hybridizations were performed using Affymetrix GeneChips Human Genome U133A [22].

All measurements have been performed in triplicate. The hybridizations were randomly assigned to 6 different groups in order to randomize the experimental effects of measuring at different points in time. The normalization of the data was performed with Rosetta Resolver Version 5 and subsequently the expression data was log transformed [23]. The biological process of interest for the dataset which is used for the analysis in this article is skeletal development.

2.2. ANOVA and PCA

ANOVA can be used in microarray data analysis to investigate the significance of the effects from factors which could possibly influence the gene expression. The ANOVA fixed-effects model [4] in which three of the possible factors of interest are incorporated is given by expression (2). In this model the measured gene expression (X_{ijkr}) is assumed to be the result of the added effects of the factors Time (T), Treatment (S) and Gene (G) and their interactions over time point i, treatment j, gene k and replicate r:

$$X_{ijkr} = \mu + T_i + S_j + G_k + TS_{ij} + TG_{ik} + GS_{jk} + TSG_{ijk} + \varepsilon_{ijkr} \quad (2)$$

with μ as the overall mean. The remaining variation is captured in the error term ε_{ijkr}. In the normal application of ANOVA, the sum of squares and mean squares are calculated for each factor and interaction, and the significance of the effects is calculated. Here, instead of calculating sums of squares, the interaction matrices are analysed with PCA [14] to identify genes with interesting biological behaviour. Because uninteresting sources of variance have been removed with ANOVA, the results are much more interpretable than a PCA applied to the original data.

The gene–treatment interaction effect can be interpreted as the response of a gene to the treatments, additional to general gene, time and treatment effects. This is of course interesting to the biologists who are looking for genes responding to the treatments incorporated in the experiment. Genes are interesting when they can be correlated with for instance a specific treatment or time point, depending on the biological question for the dataset. For the visualisation of these interactions, the technique of PCA is especially suited because it focusses on the directions of maximal variance: the interesting genes are likely to be among those with large positive or negative values.

In microarray analysis the number of measured genes is substantial; however, not all genes are expected to be involved in the biological process of interest. The aim is therefore to find a meaningful reduced set of genes. This set can consist of unknown genes as well as genes known to be involved in the process. One way to make the selection of genes from the results of the analysis with ANOVA and PCA is to use the Hotelling T^2 distribution [14]. For different values of the cutoff parameter, different numbers of genes are obtained.

2.3. Robust ANOVA

Estimates of main effects and interactions can be severely affected by outlying observations. These outliers can be caused either by measurement errors, or consist of genes showing a behaviour that is very different from the bulk of the genes, e.g. as a result of a treatment. In the latter case, these outliers are the most interesting parts of the data. Moreover, ANOVA requires normally distributed data, which in practice is not often the case. We here analyse three ways to perform more robust variants of ANOVA. The first relies on robust location estimates, and the other two on rank transformations. Combinations are possible, too. The classical, non-robust ANOVA will be indicated with the label “CL”.

2.3.1. Robust location

An alternative to the mean as a location parameter, less influenced by outliers, is the trimmed mean. When calculating the 20% trimmed mean for instance, the mean is calculated from the data which remain after the upper and lower 20% of these data are removed. The most extreme form of trimming is taking the median of the data: then, the upper and lower 50% of the data are excluded from influencing the location of the data.

Thus, when fitting the ANOVA model from Eq. (2), μ is estimated by a trimmed mean or even a median instead of the normal mean. Similarly, trimmed means or medians are used for the calculation of the main effects and interaction effects. This method will be indicated with “RL” in the following. A more refined version for estimating robust effects would be the median polish method [24]; in this paper we do not pursue this further.

2.3.2. Rank transformation

Many nonparametric methods apply a rank transformation to the data, to decrease the influence of outlying values. For ANOVA, the rank

transformation has been advocated by Iman [25]: with normally distributed data the loss of power, associated with nonparametric methods in general, is limited, but for non-normal data, the rank transformation is more powerful. Harwell and Serlin [26] suggest to use the rank transformation when the data are skewed or heavy-tailed.

In this method, indicated with "RT–CL", the original values in the whole dataset are converted to ranks. For the hMSC dataset this results in values from 1 to the number of datapoints: $22{,}283 \times 10 \times 4$. Again, the resulting data cube (now containing ranks) is analysed by ANOVA to separate sources of variance. Here, too, means can be used as location parameters, but also trimmed means or medians. Thus, the rank transformation method can be combined with the RL method. We will use means as location parameters and denote this method in the following as "RT–CL".

2.3.3. Aligned rank

An extension of the rank transformation is the aligned rank transformation method (AR), suggested by Hartlaub et al. [27]. The AR method was invented to reduce the influence of large main effects, resulting in an increase of the type I error as reported by [28]. The method was described by Gao and Song, Mansouri and Chang and even earlier by Conover and Iman [11,29,30].

The difference between the RT–CL and the AR method lies in the fact that in the AR method the main effects are subtracted from the data before the data are rank transformed. This is done to compensate for the possible differences in the size of these main effects. In literature this method has shown good performance, but the dimensions of the data were substantially smaller.

Again, this method can be combined with the RL method, even in several ways. One can use the RL method for the robust estimation of the main effects, subtract these and then do the rank transformation. After the rank transformation, one can use either means or a robust location measure. This leads to four possible combinations: CL–AR–CL, RL–AR–CL, CL–AR–RL and RL–AR–RL. In this paper, we only consider the first of these cases and refer to it with ART–CL. The other combinations perform worse, and are therefore not taken into account.

2.4. Evaluation

In comparing the different approaches, we concentrate on the main effects and the interaction between gene and treatment. We use PCA for the visualisation of the interaction terms, and compare the different interaction matrices by Procrustes analysis [17]. Using a set of 50 predefined GO terms, available in the Supplementary material, we set up ROC plots [31,20] that tell us how well each method reproduces current biological knowledge on bone formation.

2.4.1. Main effects

Robust estimates of main effects are expected to be different from the classical estimates using the mean. As a result the calculated interactions will also differ. Therefore the conclusions from the results are influenced depending on the type of robust analysis.

2.4.2. Interaction matrices

Due to the large number of genes it is difficult to identify the changes of genes in the score plots from the interaction matrices. In fact we are interested in the differences between the plots; we want to know which genes have a similar location in the plots and, even more interesting, which genes are at different locations.

A technique which can be applied to compare and investigate the topology of data points in two multivariate datasets with equal dimensions is Procrustes analysis [17]. The goal of Procrustes analysis is to find a transformation, either a translation, rotation or reflection, which results in the best match between the point configurations in both plots. Thus, Procrustes analysis can be applied to investigate the similarity of comparable data tables. An indication of the difference between the classical interaction matrix and its robust counterparts can be given for each gene, making it possible to identify the genes which show the largest difference. In the plots below, we focus – rather arbitrarily – on the twenty genes showing the largest Procrustes error (in each case). Moreover, the overall Procrustes error indicates which robust interaction matrix deviates most from the classical approach.

2.4.3. Comparison with known biology

Eventually, one is interested in finding new relations and processes, influenced by the treatments. The minimal requirement of a new method is that the processes which are expected to be found, are indeed returned by the analysis (when the experiment has been performed correctly). If that is not the case, chances are slim for finding relevant but previously unknown relations. A method to evaluate the agreement between current knowledge about a biological system and the results found with a microarray experiment is the application of ROC plots [20]. Annotation information from the Gene Ontology [18] database is used to perform the evaluation of the results. The terms associated with the Biological Function branch of GO were applied as a description of the genes which are found. Before any GO analysis was performed a list of terms expected to be found was composed by biological experts [20], available in the Supplementary material. This set is used as a reference for the results of the GO enrichment analysis, which is performed next. The terms in the list are considered as positive results and can be applied to calculate sensitivity and specificity at different cutoffs in the GO enrichment. By plotting one minus the specificity against the sensitivity, the ROC

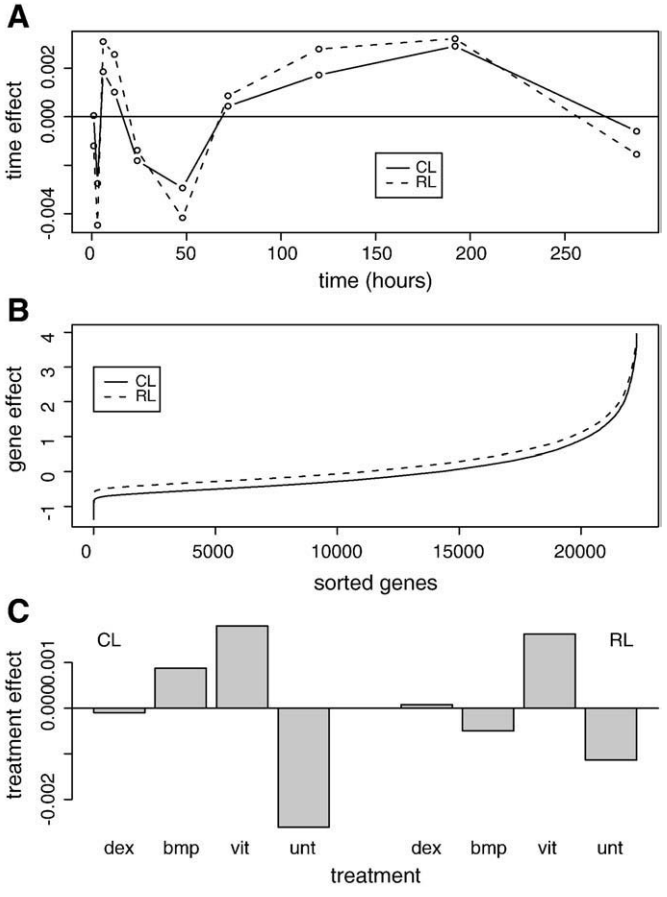

Fig. 1. The main effects of the CL and RL method of the factors time (A), gene (B) and treatment (C).

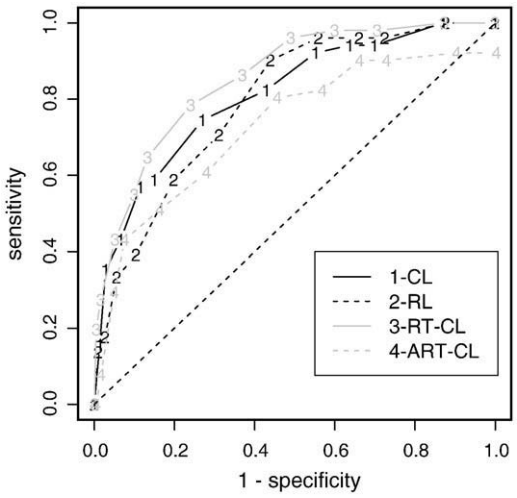

Fig. 2. ROC plot of all methods with genes selected with $\alpha = 1.10^{-2}$. The RT–CL method shows the best results, because the line is closest to optimum in the upper left corner. The ART–CL method shows the worst performance, it finds the smallest number of processes for any given cutoff for GO enrichment.

curve shows whether the results are better than results found by chance alone.

Here we use the ROC plots to compare selected genes found with robust methods and the classical application of ANOVA. In both cases, the Hotelling T^2 distribution is used to define a threshold ($\alpha = 0.01$). With this selection, enrichments are calculated for all processes in the biological process category. These results are summarised in an ROC plot.

2.4.4. Software

Calculations are performed in R [32]; the GO enrichment calculations are performed with the R package GOstats [33]. In-house scripts for the nonparametric ANOVA and the ROC curves are available upon request.

3. Results

The results for the RL method using trimmed means conform expectations: they are in between the CL results and the results of RL using medians. Therefore we will not discuss trimmed means further, and will in the remainder use medians with the RL method. First, the main effects of the CL and RL methods are shown; the main effects of the rank-based methods are comparable in their profile, but are less easily comparable with the standard main effects because they are on a much larger scale due to the rank transformation. Next, the results of the two-factor interaction gene–treatment are used to make comparisons of the different nonparametric methods and the CL method.

3.1. Main effects

The main effects for the factors time, gene and treatment are represented in Fig. 1A, B and C, respectively. From the figure it is clear that the robust and classical estimates are different. For the factor time the absolute values for the robust method are bigger. The main effect for the factor gene is approximately 0.2 larger for the RL method than for the CL method. This is the logical effect of the difference between the overall mean and overall median (5.563 and 5.337, respectively). Because the averages are taken over large numbers of data points the main effects alone are not very informative. In contrast to the effects of time and gene, the effect of the factor treatment is larger for the CL method. Obviously the differences will result in different interaction matrices as well.

3.2. Comparison with known biology

When selecting genes based on the methodology presented here one would expect certain GO categories to be significantly enriched. Among those are: ossification, skeletal development and others [20]. First, gene selections are made for each of the nonparametric methods, using the cutoff $\alpha = 1 \cdot 10^{-2}$ for the Hotelling T distribution. The agreement of the resulting gene selections with the predefined categories is shown in Fig. 2. As usual, a true positive is a gene belonging to a GO category (or one of its descendants) present in the predefined categories. The ROC curve of the RT–CL method is running closest to the optimum in the upper left corner. The ART–CL method runs below the other methods. This means that a gene selection performed with the RT–CL method gives the best balance between false negatives and false positives.

3.3. Gene–treatment interaction

The two-factor interaction of gene and treatment contains information about which genes have a specific response to the treatments. It is

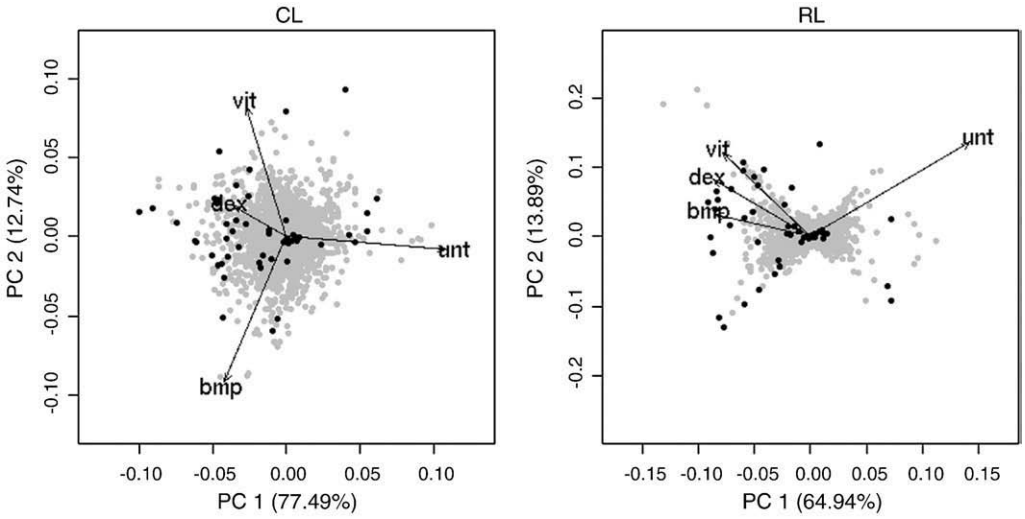

Fig. 3. Biplots of interaction matrix from interaction of gene and treatment: the CL method (left) and the RL method (right). Genes are represented with points. A group of genes with large differences between the methods, which are known to be involved in osteogenesis are indicated with black points. Loadings, representing treatments are shown with arrows.

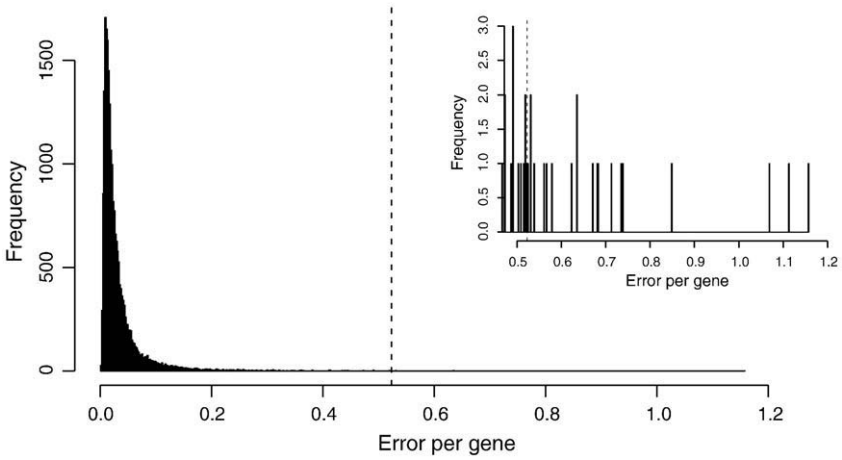

Fig. 4. Histogram generated from the Procrustes comparison between results from the CL method and the RL method. The 20 genes with the largest difference are shown in the inserted enlarged part of the right tail of the distribution.

possible to visualise the information in the gene–treatment interaction matrix with biplots, to correlate genes with specific treatments [14].

As an example of the kind of differences one can expect when comparing classical and robust methods, the biplots from the gene–treatment interaction matrices of the CL and RL methods are shown in Fig. 3. The percentage of explained variance on the first PC is smaller in the robust case. The loadings still show the separation between the three treatments and the untreated control in the first PC. However, there seems to be less separation between the three osteogenic treatments in the second PC of the RL method. In the plots a group of genes is indicated for which the individual Procrustes error is largest (see below).

For all methods a separation between the untreated and osteogenic treatments can be seen in the loadings, but the shape of the distribution of genes and the distance of genes to the center depend on the method. To see which methods are similar, we perform Procrustes analysis.

First, a comparison is made between the CL and RL methods. The RL interaction matrix is rotated and scaled to find the best match with the CL interaction matrix. As a result of the analysis the rotated RL interaction matrix can be compared to the original CL interaction matrix. First, the distribution of the differences between the rotated and the original is depicted in Fig. 4. Most errors are relatively small compared to the outliers in the right tail of the distribution. This means that the relative orientation of most genes has not changed, because the bulk is in the center of the biplots, relatively irresponsive to the effects of the treatments. One would then expect that these elements experience only small changes when comparing the two methods.

Second, from the Procrustes analysis the overall Procrustes error can be calculated by taking the sum of squared differences between the original CL data matrix and the rotated RL matrix. The overall Procrustes error is shown in Table 1. Perhaps not surprisingly the rank-based methods show a larger difference with the CL method than the RL method.

Table 1
Comparison of the nonparametric methods RL, RT-CL and ART-CL with the classical ANOVA-PCA.

RL vs CL	71.2
RT–CL vs CL	96.0
ART–CL vs CL	96.4

The overall Procrustes error is calculated as the sum of squared differences between the original matrix and the rotated target matrix.

Third, to investigate which genes have a large Procrustes difference, when compared with the classical ANOVA–PCA, we calculate the difference between the scores on PCs 1 and 2 of the original interaction matrix and the optimal match with the corresponding scores of the interaction matrix from the nonparametric method. The difference matrix, obtained after this Procrustes rotation, is then inspected with PCA. In Fig. 5 the scores corresponding to the genes are shown for the methods RL, RT–CL and ART–CL. In the figures a number of genes are indicated in black for which it is known that they are involved in osteogenesis (from literature) and which have a large Procrustes difference. Note that these genes are different from the genes highlighted in Fig. 3. There are 7 genes in the set which are found by the three methods; OMD, LEPR, IGFBP2, GAS1, COL11A1, BGLAP and ADAM12. More detailed information and references for these and other genes mentioned here, can be found in the Supplementary material. The set of genes MMP7, LEP, IBSP, FRZB, CHI3L1 and CHRDL1 are present among the set of genes with the largest difference for the methods RL and ART–CL. There are three genes which have a large error for the RL method; BMP6, LIF and PER1. Finally, the difference for genes HEY1 and ID3 is apparent for the RT–CL and the ART–CL methods.

The p-value cutoff for the genes ranked based on the Hotelling T^2 distribution can also be used when we look specifically at the gene–treatment interaction term. The difference between the methods is difficult to determine because there is a lot of overlap between the interesting genes which are found by the methods. There is a difference between the methods though, and it is in the distribution of the interesting genes in the selection. To describe these differences, we have chosen to use the 100 most extreme genes from the biplot. These are the genes which are most influenced by the treatments in the experiment. From the selection of 100 genes for each method we will discuss a limited number of genes known to be involved in osteogenesis based on literature. The genes BMP6 and CDKN1C are found specifically with the RT–CL method. Furthermore the genes PER1, HEY1 and PTGER2 are only present in both rank-based methods (RT–CL and ART–CL) whereas COL11A1, CHRDL1, ID3 and IGFBP2 are discovered with the classical and RL method.

4. Discussion and conclusion

As explained before in the Introduction, the estimates for the main effects and the interactions are influenced by outlying observations. These can be the result of real biological differences in gene behaviour, or result from measurement errors. We have shown that taking a nonparametric approach which does not rely on distributional

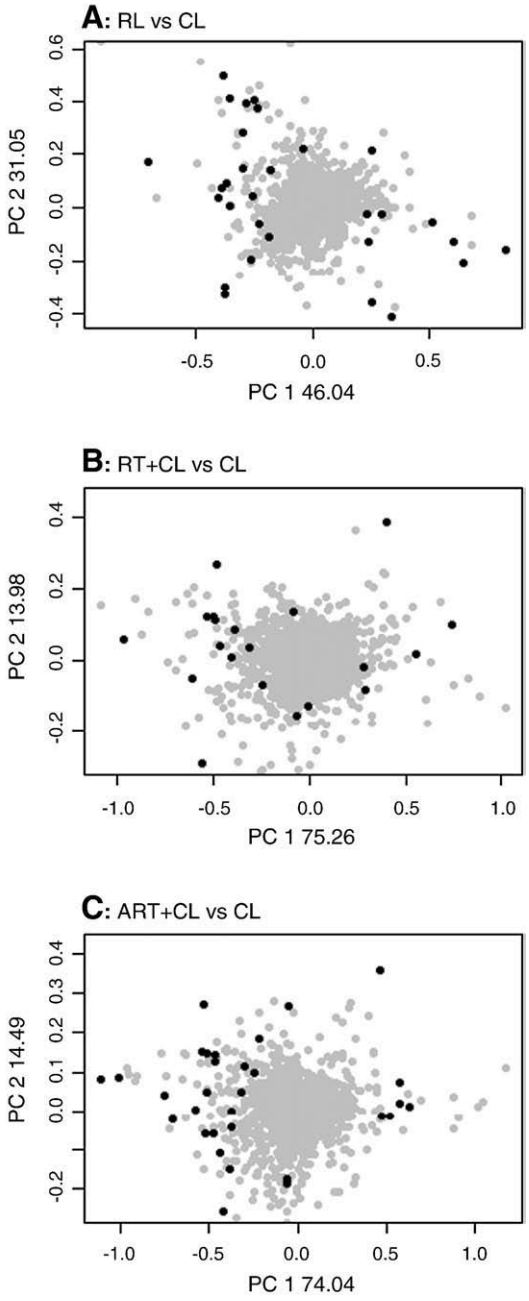

Fig. 5. Results of PCA performed on the difference between the original X matrix and the rotated Y matrix. Genes with a large difference will be on the outside of the plot. Comparisons were made with the classical ANOVA–PCA as reference. It was then compared with the RL (A), RT–CL (B) and the ART–CL (C) and methods, respectively. Each point represents a gene. Several genes with low p-values, known from literature to be responsive to osteogenic treatments, are shown with black points. The threshold for the p-values is $p<1\cdot10^{-5}$ for the Hotelling T^2 distribution on the first 2PCs.

assumptions of the data gives different estimates of main effects and interactions, which are not influenced by outliers.

Robust ANOVA methods lead to quite different results, compared to classical ANOVA, for the data in this paper. This is an indication that the assumptions of parametric ANOVA are not justified for this dataset; otherwise, the differences would be less substantial. Application of robust ANOVA leads to different estimates of main effects and interactions. As a result, different biological conclusions will be drawn, e.g., different genes will be selected on the basis of the gene–treatment interaction. Furthermore, the robust methods take slightly different approaches: using the median instead of the mean is different from using ranks instead of expression values. The sizes of the differences may depend on the data at hand.

When selecting genes with an interesting gene–treatment interaction, some genes are found with one method, and not with the other. The RT–CL method, which is shown to give the best representation of known biology, finds the genes BMP6 and CDKN1C for instance. These genes are not present in the selections made with the other methods. To give a more general picture of the biological results, ROC plots are drawn. For this dataset the results from the ROC plots for expected GO categories indicate that the rank transformation method with normal means gives results which are most in agreement with current biological knowledge of the system. From the results it can be concluded that it is not advisable to use the ART–CL method.

In addition to the methods that perform robust ANOVA shown in the Results section, we have applied several other robust methods as well. The rank transformation based methods can also be applied with the robust location estimate. The results from these, RT–RL and ART–RL, were much worse than the other methods and are therefore not presented here. We also evaluated the interaction matrices (obtained from either robust or classical ANOVA) using robust PCA (ROBPCA) [34]. The goal of robust PCA is to obtain principal components that are not influenced much by outliers in the data. Applying robust PCA on the interaction matrices showed no real differences with normal PCA, therefore results are not shown. A final opportunity for further robust analysis is the Procrustes rotation: by rotating the data in such a way that the majority of the points is correctly aligned in the two datasets, the differences of the remaining points may stand out more clearly. We did not pursue this avenue further, since the standard Procrustes analysis provided enough insight already.

Acknowledgements

The first author would like to thank Schering-Plough for financial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chemolab.2009.04.011.

References

[1] J.D. Boyce, I. Wilkie, M. Harper, M.L. Paustian, V. Kapur, B. Adler, Genomic scale analysis of *Pasteurella multocida* gene expression during growth within natural chicken host, Infection and Immunity 70 (2002) 6871–6879.
[2] S. Davletova, L. Rizhsly, L. Lian, Z. Shengqiang, D.J. Oliver, J. Coutu, V. Shulaev, K. Schlauch, R. Mittlera, Cytosolic ascorbate paroxidase 1 is a central component of the reactive oxygen gene network of *Arabidopsis*, The Plant Cell 17 (2005) 268–281.
[3] S. Raman, R. Hazra, C.C. Dasher, R.N. Husson, Transcription regulation by the mycobacterium tuberculosis alternative sigma factor sigD and its role in virulence, Journal of Bacteriology 189 (2004) 6605–6616.
[4] M.K. Kerr, M. Martin, G.A. Churchill, Analysis of variance for gene expression microarray data, Journal of Computational Biology 7 (2000) 819–837.
[5] L.W. Lee, B.D. Whitmore, Models for microarray gene expression data, Journal of Biopharmaceutical Statistics 12 (2002) 1–19.
[6] R.D. Wolfinger, G. Gibson, E.D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, R.S. Paules, Assessing gene significance from cDNA microarray expression data via mixed models, Journal of Computational Biology 8 (2001) 625–637.
[7] B.P. Durbin, J.S. Hardin, D.M. Hawkins, D.M. Rocke, A variance-stabilizing transformation for gene-expression microarray data, Bioinformatics 18 (2002) S105–S110.
[8] X. Cui, M.K. Kerr, G.A. Churchill, Transformations for cDNA microarray data, Statistical Applications in Genetics and Molecular Biology, 2.1, 2007.
[9] S.C. Geller, J.P. Gregg, P. Hagerman, D.M. Rocke, Transformation and normalization of oligonucleotide microarray data, Bioinformatics 19 (2003) 1817–1823.
[10] W. Huber, A. von Heydebreck, H. Sultmann, A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression, Bioinformatics 18 (2002) S96–S104.
[11] X. Gao, P.X.K. Song, Nonparametric tests for differential gene expression and interaction effects in multi-factorial microarray experiments, BMC Bioinformatics 6 (2005).

[12] Y. Zhao, W. Pan, Modified nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments, Bioinformatics 19 (2003) 1046–1054.
[13] O.G. Troyanskaya, M.E. Barber, P.O. Brown, D. Botstein, R.B. Altman, Nonparametric methods for identifying differentially expressed genes in microarray data, Bioinformatics 18 (2002) 1454–1461.
[14] J.R. de Haan, R. Wehrens, S. Bauerschmidt, E. Piek, R.C. van Schaik, L.M.C. Buydens, Interpretation of ANOVA models for microarray data using PCA, Bioinformatics 23 (2007) 184–190.
[15] H.F. Gollob, A statistical model which combines features of factor analytic and analysis of variance techniques, Psychometrika 33 (1968) 73–114.
[16] R.W. Zobel, M.J. Wright, H.G. Gauch, Statistical analysis of a yield trial, Agronomy Journal 80 (1988) 388–393.
[17] J.C. Gower, Generalized Procrustes analysis, Psychometrika 40 (1975) 33–51.
[18] The Gene Ontology Consortium. Gene ontology: tool for the unification of biology, Nature Genetics 25 (2000) 25–29.
[19] P. Khatri, S. Draghici, Ontological analysis of gene expression data: current tools, limitations and open problems, Bioinformatics 21 (2005) 3587–3595.
[20] J.R. de Haan, E. Piek, R. Wehrens, R.C. van Schaik, S. Bauerschmidt, L.M.C. Buydens, Beyond single *p*-value cutoffs: Methods to improve decision making in go enrichment analysis of microarray experiments, submitted for publication.
[21] N.R. Jørgensen, Z. Henriksen, O.H. Sørensen, R. Civitellis, Dexamethasone, BMP-2, and 1,25-dihydroxyvitamin D enhance a more differentiated osteoblast phenotype: validation of an in vitro model for human bone marrow-derived primary osteoblasts, Steroids 69 (2004) 219–226.
[22] R.J. Lipshutz, S.P.A. Fodor, T.R. Gingeras, D.J. Lockhart, High density synthetic oligonucleotide arrays, Nature Genetics 21 (1999) 20–24.
[23] L. Weng, H. Dai, Y. Zhan, Y. He, S.B. Stepaniants, D.E. Bassett, Rosetta error model for gene expression analysis, Bioinformatics 22 (2006) 1111–1121.
[24] F. Mosteller, J. Tukey, Data Analysis and Regression, Addison-Wesley, Reading, MA, 1997.
[25] R.L. Iman, A power study of a rank transform for the two-way classification model when interactions may be present, Canadian Journal of Statistics 2 (1974) 227–239.
[26] M.R. Harwell, R.C. Serlin, A nonparametric test statistic for the general linear model, Journal of Educational Statistics 14 (1989) 351–371.
[27] B.A. Hartlaub, A.M. Dean, D.A. Wolfe, Rank-based test procedures for interaction in the two-way layout with one observation per cell, Canadian Journal of Statistics 27 (2000) 863–874.
[28] R.C. Blair, S.S. Sawilowsky, J.J. Higgens, Limitations of the rank transform statistic in tests for interactions, Statistical Science 9 (1987) 94–108.
[29] H. Mansouri, G.H. Chang, Rank tests for interaction, Computational Statistics & Data Analysis 17 (1993) 417–426.
[30] W.J. Conover, R.L. Iman, Rank transformations as a bridge between parametric and nonparametric statistics, The American Statistician 35 (1981) 124–129.
[31] C.D. Brown, H.T. Davis, Receiver operating characteristics curves and related decision measures: a tutorial, Chemometrics and Intelligent Laboratory Systems 80 (2006) 24–38.
[32] R Development Core Team. R, A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0, 2006.
[33] S. Falcon, R. Gentleman, Using GOstats to test gene lists for GO term association, Bioinformatics 23 (2007) 257–258.
[34] M. Hubert, S. Engelen, Robust PCA and classification in the biosciences, Bioinformatics 20 (2004) 1728–1736.