Global testing of differential gene expression

Ulrich Mansmann   Reinhard Meister   Manuela Hummel

April 25, 2006

Contents

1 Abstract

In studies about differential gene expression between different clinical diagnoses the main interest may often not be in single genes but rather in groups of genes that are associated with a pathway or have a common location in the genome. In such cases it may be better to perform a global test because the problems of multiple testing can be avoided. The approach presented here is an ANCOVA global test on phenotype main effect and gene–phenotype interaction.

Testing many pathways simultaneously is also possible. This, of course, causes again need for correction for multiple testing. Besides the standard approaches for correction we introduce a closed testing procedure in which the experiment–wise error rate equals the required level of confidence of the overall test.

This document was created using R version 2.3.0 and versions 2.2.0 and 4.2.0 of the packages *GlobalAncova* and *globaltest* respectively.

2 Changes in Version 2.x.x

- The major modification in the new version is the transfer from simple two group comparisons to a general linear model framework where arbitrary clinical variables (in especially with more groups or also continuous ones), time trends, gene–gene interactions, co–expression and so forth can be analysed.

- According to the new framework also the diagnostic plots are more flexible. The variable defining the coloring of bars can now be specified by the user, see section 6 for details.

- A bug was fixed concerning testing only a single gene for differential expression with the global ANCOVA F–test.

- Within the closed testing procedure a bug was fixed concerning testing non–disjunct groups of genes.

3 Introduction

The ANCOVA global test is a test for the association between expression values and clinical entities. The test is carried out by comparison of linear models via the extra sum of squares principle. If the mean expression level for at least one gene differs between corresponding models the global null hypothesis, which is the intersection of all null hypotheses for the single genes, is violated. As our test is based on the sum of gene-wise reduction in sum of squares due to phenotype, all systematic differences in gene expression between phenotypes equally contribute to the power of the test.

Single genes are not, in general, the primary focus of gene expression experiments. The researcher might be more interested in relevant pathways, functional sets or genomic regions consisting of several genes. Most of the current methods for studying pathways analyse differential expression of single genes. In these methods pathways where many genes show minor changes in their expression

values may not be identified. Goeman's global test and the ANCOVA global test were designed to address this issue.

Applying global tests for differential expression in pathways substantially reduces the number of tests compared to gene-wise multiple testing. The amount of correction for multiple testing decreases. Function (KEGG, GO) or location (chromosome, cytoband) could be used as grouping criteria, for example.

We want to compare our method with the global test of Goeman et al., 2004 [1]. Therefore text and examples in this document follow to a certain extent the vignette presented in the R-package *globaltest*. Our function `GlobalAncova` tests whether the expectation of expression levels differs between biological entities for a given group of genes. This vignette has its focus on the practical use of the test. For more details about the mathematical background and the interpretation of results, we refer to the paper by Mansmann and Meister, 2005 [4].

This document shows the functionality of the R-package *GlobalAncova*. The datasets, all necessary R–packages and our package *GlobalAncova* are available from the Bioconductor website (*www.bioconductor.org*).

First we load the packages and data we will use.

```
> library(GlobalAncova)
> library(globaltest)
> library(golubEsets)
> library(hu6800)
> library(vsn)
> library(multtest)
> data(golubMerge)
> golubM <- update2MIAME(golubMerge)
> golubX <- vsn(golubM)
```

This creates a dataset `golubX`, which is of the format *exprSet*, the standard format for gene expression data in BioConductor. It consists of 7129 genes and 72 samples (the data are from [2]). We used *vsn* to normalize the data. Any other normalization method may be used instead. From several phenotype variables we use "ALL.AML" as the clinical diagnoses of interest. ALL and AML are two types of acute leukemia. There are 47 patients with ALL and 25 with AML.

4 Global Testing of a Single Pathway

4.1 Golub Data and Cell Cycle Pathway

Suppose we are interested in testing whether AML and ALL have different gene expression patterns for certain pathways, for example from the KEGG database.

All Genes
We start by applying our test to all genes in the Golub dataset so that differences in the overall gene-expression pattern can be demonstrated.

```
> gr <- as.numeric(golubX$ALL.AML == "ALL")
> ga.all <- GlobalAncova(xx = exprs(golubX), group = gr,
+     covars = NULL, perm = 100, test.genes = NULL)
```

The first input *xx* is a 7129 × 72 matrix that contains the expression values of all genes and samples. The second input *group* is a vector that defines the clinical diagnosis for the 72 patients.

Note that the new version of *GlobalAncova* is not any more restricted to the analysis of dichotomous phenotype groups. More complex tasks like variables with more groups or also continuous ones, time trends, gene–gene interactions and co–expression can be performed as well. Some examples will be given in section 4.2. The realization of such tasks is done by definition of two linear models that shall be compared via the extra sum of squares principle. Hence model formulas for the full model containing all parameters and the reduced model, where the terms of interest are omitted, have to be given. An alternative is to provide the formula for the full model and a character vector naming the terms of interest. Consequently we could run the same analysis as above with two possible further function calls shown below (output is omitted). In both cases a data frame with information about all variables for each sample is required. In the case of microarray data this can be the corresponding `pData` object.

```
> GlobalAncova(xx = exprs(golubX), formula.full = ~ALL.AML,
+     formula.red = ~1, model.dat = pData(golubX), perm = 100)
> GlobalAncova(xx = exprs(golubX), formula.full = ~ALL.AML,
+     test.terms = "ALL.AML", model.dat = pData(golubX),
+     perm = 100)
```

To avoid alpha–inflation due to correlated data and effects of non–normality of the data tests for significance of the resulting F–ratios are performed using a permutation test approach. The argument *perm* defines the number of permutations. The number of permutations is 10,000 for default but in the current version the test for many genes can take quite a long time with 10,000 permutations. Here we set *perm* to just 100 or 1000 so that creating this vignette will not last too long. For getting more reliable results one should recompute the examples with more permutations. We are currently looking for a reliable asymptotic test as a chance to get around with too time consuming permutations.

The result is a typical ANOVA table with information about sums of squares, degrees of freedom and mean sums of squares for the effect and error term, respectively. Besides F–statistics and classical F–test p–values, there are also p–values from the permutation test. The names of all involved parameters are displayed as well as the name(s) of the tested effect(s).

```
> ga.all

$effect
[1] "gr"

$ANOVA
            SSQ      DF         MS
Effect  14800.04    7129  2.0760336
Error  182564.22  499030  0.3658382

$test.result
```

```
F.value 5.674732
p.value 0.000000
p.perm  0.000000

$terms
[1] "(Intercept)" "gr"
```

From this result we conclude that the overall gene expression profile for all 7129 genes is associated with the clinical outcome. This means that samples with different AML/ALL status tend to have different expression profiles. We expect most pathways (especially the ones containing many genes) also to be associated with the phenotype groups.

If we apply Goeman's global test we get

```
> gt.all <- globaltest(golubX, "ALL.AML")
> gt.all

Global Test result:
Data: 72 samples with 7129 genes; 1 pathway tested
Model: logistic
Method: Asymptotic distribution

    Genes Tested Statistic Q Expected Q sd of Q   P-value
all  7129   7129     53.992         10  2.5554 1.5133e-10
```

Both tests show that the data contain overwhelming evidence for differential gene expression between AML and ALL.

Cell Cycle Pathway

Now we ask the more specific question of whether there is evidence for differential gene expression between both diagnoses restricted to genes belonging to the cell cycle pathway. First we load all KEGG pathways.

```
> kegg <- as.list(hu6800PATH2PROBE)
```

The list kegg consists of 172 pathways. Each pathway is represented by a vector of gene names. We are mainly interested in the cell cycle pathway which has the identifier "04110" in the KEGG database. It corresponds to 106 probe sets on the hu6800 chip.

```
> cellcycle <- kegg[["04110"]]
```

We apply the global test to this pathway using the option *test.genes*.

```
> ga.cc <- GlobalAncova(xx = exprs(golubX), group = gr,
+     test.genes = cellcycle, perm = 1000)
> ga.cc

$effect
[1] "gr"

$ANOVA
```

```
            SSQ   DF        MS
Effect   271.6363  106 2.5626067
Error   2574.4204 7420 0.3469569

$test.result

F.value 7.38595
p.value 0.00000
p.perm  0.00000

$terms
[1] "(Intercept)" "gr"
```

Also with *globaltest* we get a very small p–value

```
> gt.cc <- globaltest(X = golubX, Y = "ALL.AML", genesets = cellcycle)
> gt.cc

Global Test result:
Data: 72 samples with 7129 genes; 1 pathway tested
Model: logistic
Method: Asymptotic distribution

      Genes Tested Statistic Q Expected Q sd of Q   P-value
[1,]   106    106       66.646     9.6983  3.4502 1.7653e-08
```

The test results clearly indicate that the expression pattern of the cell cycle pathway is different between the two clinical groups.

Adjusting for Covariates

Covariate information can be incorporated by specifying the *covars* option. For example if we want to adjust for Source, the institution that provided the samples, we can do this by

```
> ga.cc.source <- GlobalAncova(xx = exprs(golubX), group = gr,
+     covars = golubX$Source, test.genes = cellcycle, perm = 1000)
> ga.cc.source

$effect
[1] "gr"

$ANOVA
             SSQ   DF        MS
Effect    49.60491  106 0.4679709
Error   2399.95706 7102 0.3379269

$test.result

F.value 1.384828614
p.value 0.005760897
p.perm  0.083000000
```

```
$terms
[1] "(Intercept)"          "gr"
[3] "golubX$SourceCCG"     "golubX$SourceDFCI"
[5] "golubX$SourceSt-Jude"
```

With the more general function call we would simply adjust the definitions of model formulas, namely $formula.full = \sim ALL.AML + Source$ and $formula.red = \sim Source$.

The source of the samples apparently has some explanatory effect on the outcome resulting in a smaller F–ratio than in the model without adjusting.

With the *globaltest* we get a higher p–value.

```
> gt.cc.source <- globaltest(X = golubX, Y = ALL.AML ~
+     Source, genesets = cellcycle)
> gt.cc.source

Global Test result:
Data: 72 samples with 7129 genes; 1 pathway tested
Model: logistic, ALL.AML ~ Source
Adjusted: 11.5 % of variance of Y remains after adjustment
Method: Asymptotic distribution

     Genes Tested Statistic Q Expected Q sd of Q P-value
[1,]   106    106      15.782     10.468  6.0623 0.16549
```

Permutation based p–values can also be obtained with Goeman's test, however only when covariates are absent.

4.2 van t'Veer Data and p53–Signalling Pathway

We present another example from a study on breast cancer from [6]. This example illustrates how more complex tasks than comparing just two clinical groups can be performed with *GlobalAncova*. A subset of the data consisting of the expression values for 96 patients without *BRCA1* or *BRCA2* mutations is available with the package. The dataset (vantVeer) is restricted to 1113 genes associated with 9 cancer related pathways that are provided as a list named (pathways), too. Information about some of the originally surveyed covariates is stored in phenodata. The tumour suppressor protein p53 contributes as a transcription factor to cell cycle arrest and apoptosis induction. Therefore, first the p53-signalling pathway is selected as a candidate, where differential expression between relevant prognostic groups, defined by the development of distant metastases within five years, was expected.

```
> data(vantVeer)
> data(phenodata)
> data(pathways)
> metastases <- phenodata$metastases
> p53 <- pathways$p53_signalling
```

We get a significant result with the global ANCOVA.

```
> set.seed(111)
> vV.1 <- GlobalAncova(xx = vantVeer, group = metastases,
+     test.genes = p53, perm = 1000)
> vV.1

$effect
[1] "metastases"


$ANOVA
            SSQ   DF         MS
Effect  2.893417   33 0.08767929
Error  97.424573 3102 0.03140702


$test.result


F.value 2.791710e+00
p.value 2.307974e-07
p.perm  7.000000e-03


$terms
[1] "(Intercept)" "metastases"
```

The test result illustrates that the theoretical p–value is probably over–optimistic.

Analysis of Arbitrary Clinical Variables

In the new version of the package also clinical variables with more than two groups can be considered. For demonstration we investigate differential expression for the three ordered levels of tumor grade.

```
> set.seed(111)
> vV.3 <- GlobalAncova(xx = vantVeer, formula.full = ~ordered(grade),
+     formula.red = ~1, model.dat = phenodata, test.genes = p53,
+     perm = 1000)
> vV.3

$effect
[1] "ordered(grade).L" "ordered(grade).Q"


$ANOVA
            SSQ   DF         MS
Effect  3.638565   66 0.05512977
Error  96.679425 3069 0.03150193


$test.result


F.value 1.7500442211
p.value 0.0001979060
p.perm  0.0450000000


$terms
[1] "(Intercept)"      "ordered(grade).L" "ordered(grade).Q"
```

Gene–Gene Interaction

Now we want to go into the matter of other interesting biological questions. For example one might ask if there exists interaction between the expression of genes which the authors in [6] presented as signature for prediction of cancer recurrence and the expression of genes in a certain pathway. This question can be answered by viewing the expression values of the signature genes as linear regressors and to test their effects on the expression pattern of the pathway genes. For demonstration we pick the signature gene "AL137718" and again the p53–signalling pathway. Assume that we also want to adjust for the Estrogen receptor status. The analysis can be carried out in the following way.

```
> signature.gene <- "AL137718"
> model <- data.frame(phenodata, signature.gene = vantVeer[signature.gene,
+     ])
> set.seed(111)
> vV.4 <- GlobalAncova(xx = vantVeer, formula.full = ~signature.gene +
+     ERstatus, formula.red = ~ERstatus, model.dat = model,
+     test.genes = p53, perm = 1000)
> vV.4

$effect
[1] "signature.gene"


$ANOVA
            SSQ   DF          MS
Effect   2.667014   33 0.08081859
Error   89.867452 3069 0.02928232


$test.result

F.value 2.759979e+00
p.value 3.258265e-07
p.perm  1.100000e-02


$terms
[1] "(Intercept)"    "signature.gene" "ERstatuspos"
```

Assuming a significance level of 0.05 we get a significant effect of the signature gene on the p53–signalling pathway.

Co–Expression

Next we want to analyse co–expression regarding the clinical outcome of building distant metastases within five years. This can be done by simply adding the variable metastases to the full and reduced model, respectively. Such layout corresponds to testing the linear effect of the signature gene stratified not only by Estrogen receptor status but also by metastases.

```
> set.seed(111)
> vV.5 <- GlobalAncova(xx = vantVeer, formula.full = ~metastases +
+     signature.gene + ERstatus, formula.red = ~metastases +
+     ERstatus, model.dat = model, test.genes = p53, perm = 1000)
> vV.5
```

9

```
$effect
[1] "signature.gene"


$ANOVA
             SSQ   DF          MS
Effect  2.284391   33  0.06922396
Error  87.463681 3036  0.02880885


$test.result


F.value 2.402871e+00
p.value 1.340864e-05
p.perm  1.400000e-02


$terms
[1] "(Intercept)"    "metastases"     "signature.gene"
[4] "ERstatuspos"
```

Again we get a significant result.

Supposably the most interesting question in this case concerns differential co–expression. Differential co–expression is on hand if the effect of the signature gene behaves different in both `metastases` groups. In a one dimensional context this would become manifest by different slopes of the regression lines. Hence what we have to test is the interaction between `metastases` and `signature.gene`.

```
> set.seed(111)
> vV.6 <- GlobalAncova(xx = vantVeer, formula.full = ~metastases *
+     signature.gene + ERstatus, formula.red = ~metastases +
+     signature.gene + ERstatus, model.dat = model, test.genes = p53,
+     perm = 1000)
> vV.6

$effect
[1] "metastases:signature.gene"


$ANOVA
             SSQ   DF          MS
Effect  2.520643   33  0.07638311
Error  84.943038 3003  0.02828606


$test.result


F.value 2.700380e+00
p.value 6.197318e-07
p.perm  5.000000e-03


$terms
[1] "(Intercept)"               "metastases"
[3] "signature.gene"            "ERstatuspos"
[5] "metastases:signature.gene"
```

We observe a significant differential co-expression between the chosen signature gene and the p53–signalling pathway.

5 Testing Several Pathways Simultaneously

Systems biology involves the study of mechanisms underlying complex biological processes as integrated systems of many diverse interacting components, often referred to as pathways.

We regard the possibility to investigate differential gene expression simultaneously for several of those pathways as a contribution towards understanding biological relevant relations.

The user can apply `GlobalAncova` to compute p–values for a couple of pathways with one call by specifying the *test.genes* option. The members of each pathway to be tested must belong to genes in the expression–matrix. Afterwards a suitable correction for multiple testing has to be applied. An alternative based on the closed testing approach is described later.

Suppose for example that for sake of simplicity we want to test the first four of the cancer related pathways with the van t'Veer data. We proceed as follows.

```
> metastases <- phenodata$metastases
> ga.pw <- GlobalAncova(xx = vantVeer, group = metastases,
+     test.genes = pathways[1:4], perm = 1000)
> ga.pw
```

	genes	F.value	p.value	p.perm
androgen_receptor_signaling	72	2.3898	0.0000	0.007
apoptosis	187	1.9685	0.0000	0.012
cell_cycle_control	31	4.6399	0.0000	0.000
notch_delta_signalling	34	1.4972	0.0323	0.100

The result is a matrix whose rows correspond to the different pathways.

With the `globaltest` we get a similar matrix.

```
> gt.pw <- globaltest(X = vantVeer, Y = metastases, test.genes = pathways[1:4])
> gt.pw
```

```
Global Test result:
Data: 96 samples with 1113 genes; 4 pathways tested
Model: logistic
Method: Asymptotic distribution
```

	Genes	Tested	Statistic Q	Expected Q
androgen_receptor_signaling	72	72	22.197	9.3260
apoptosis	187	187	16.454	8.3560
cell_cycle_control	31	31	51.241	11.3470
notch_delta_signalling	34	34	12.711	8.4455

	sd of Q	P-value
androgen_receptor_signaling	3.8650	0.01059600
apoptosis	2.8665	0.01664700
cell_cycle_control	5.7933	0.00036812
notch_delta_signalling	4.1647	0.13406000

5.1 Simultaneous Adjustment of p–values

Next we show how to extract p–values for correction for multiple testing. Note however that due to the extremely high correlations between these tests, many procedures that correct for multiple testing here are inappropriate. An appropriate way of adjusting would be for example the method of Holm, 1979 [3]. An alternative to such adjustments that is not affected by correlations between tests is a closed testing procedure. For this approach you need a family of null hypotheses that is closed under intersection. Then a single hypothesis can be rejected at level α if it is rejected along with all hypotheses included in it ([5]).

For the adjustment according to Bonferroni and Holm we build a vector of the raw p–values. The function `mt.rawp2adjp` provides several adjusting methods. We here display only the raw and "Holm" adjusted p–values. To obtain the original order of the pathways we order the result of `mt.rawp2adjp` according to `index`.

```
> ga.pw.raw <- ga.pw[, "p.perm"]
> ga.pw.adj <- mt.rawp2adjp(ga.pw.raw)
> ga.result <- ga.pw.adj$adjp[order(ga.pw.adj$index), c("rawp",
+     "Holm")]
> rownames(ga.result) <- names(pathways)[1:4]
> ga.result

                          rawp  Holm
androgen_receptor_signaling 0.007 0.021
apoptosis                   0.012 0.024
cell_cycle_control          0.000 0.000
notch_delta_signalling      0.100 0.100

> gt.pw.raw <- gt.pw@res[, "P-value"]
> gt.pw.adj <- mt.rawp2adjp(gt.pw.raw)
> gt.result <- gt.pw.adj$adjp[order(gt.pw.adj$index), c("rawp",
+     "Holm")]
> rownames(gt.result) <- names(pathways)[1:4]
> gt.result

                          rawp          Holm
androgen_receptor_signaling 0.0105958152 0.031787446
apoptosis                   0.0166473874 0.033294775
cell_cycle_control          0.0003681155 0.001472462
notch_delta_signalling      0.1340573473 0.134057347
```

Allowing a family–wise error rate of 0.05 all but one pathways remain significant for both methods.

5.2 Closed Testing Procedure

Closed testing procedures ([5]) offer a versatile and powerful approach to the multiple testing problem. Implementation is non–trivial, therefore, the program given in this version should be regarded as a prototype.

In order to apply the closed testing procedure we first have to create the required family of hypotheses by building all intersections between the "natural"

hypotheses tested above and all intersections of those new hypotheses and so on.

The resulting family of hypotheses can be illustrated in a directed graph. If we just for the sake of illustration assume that we have only four hypotheses named "1", ... "4" then the node "1-2-3-4" for example stands for the global hypothesis that the genes of all four pathways are not differentially expressed. Now the interesting hypothesis "1" for example can be rejected if also the hypotheses "1-2-3-4", "1-2-3", "1-2-4", "1-3-4", "2-3-4", "1-2", ..., "1-4" are rejected. These relationships are represented by the edges of the graph.

```
Loading required package: Rgraphviz
Loading required package: graph
Loading required package: Ruuid
```
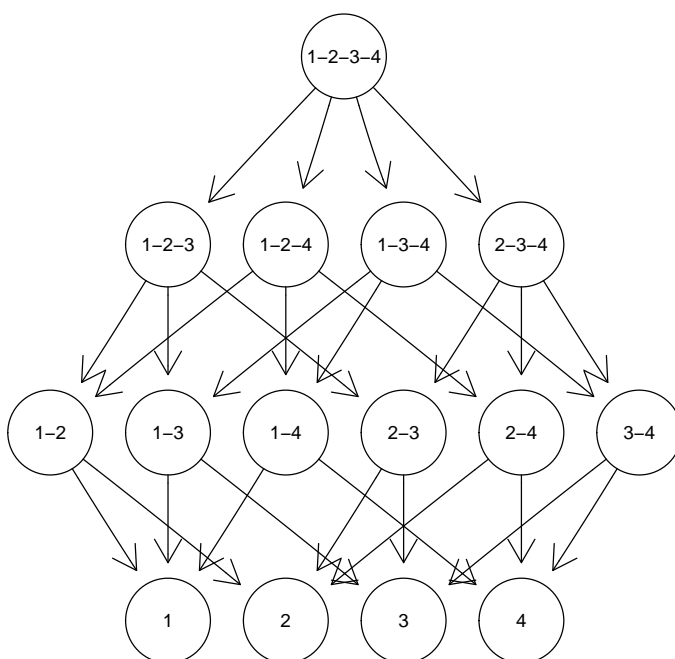

We can compute the closed testing procedure using the function

```
> ga.closed <- GlobalAncova.closed(xx = vantVeer, group = metastases,
+     test.genes = pathways[1:4], previous.test = ga.pw,
+     level = 0.05, perm = 100)
```

where *test.genes* is again a list of pathways. In order to shorten computing time we can provide the results of the previous application of `GlobalAncova` for the pathways of interest. The option *level* allows to manipulate the level of significance. *perm* again gives the desired number of permutations used in the permutation test.

Also for `GlobalAncova.closed` all three different function calls as for `GlobalAncova` itself are possible.

The function `GlobalAncova.closed` provides the formed null hypotheses (this means lists of genes to be tested simultaneously), the test results for each pathway of interest and the names of significant and non significant pathways. Names for the intersections of hypotheses are built by simply coercing the names of the respective pathways. If for a pathway one single hypothesis can not be rejected there is no need to test all the remaining hypotheses. That is why in test results of non significant pathways lines are filled with NA's after a p–value $> \alpha$ occured. Here only test results for the first pathway are displayed.

```
> names(ga.closed)

[1] "new.data"        "test.results"    "significant"
[4] "not.significant"

> rownames(ga.closed$test.results[[1]])

[1] "androgen_receptor_signaling"
[2] "androgen_receptor_signaling.apoptosis"
[3] "androgen_receptor_signaling.cell_cycle_control"
[4] "androgen_receptor_signaling.notch_delta_signalling"
[5] "apoptosis.androgen_receptor_signaling.cell_cycle_control"
[6] "apoptosis.androgen_receptor_signaling.notch_delta_signalling"
[7] "cell_cycle_control.androgen_receptor_signaling.notch_delta_signalling"
[8] "cell_cycle_control.apoptosis.androgen_receptor_signaling.notch_delta_signalling"

> rownames(ga.closed$test.results[[1]]) <- NULL
> ga.closed$test.results[1]

$androgen_receptor_signaling
     genes F.value p.value p.perm
[1,]    72  2.3898       0  0.007
[2,]   258  2.0961       0  0.010
[3,]   100  3.0409       0  0.010
[4,]   106  2.1207       0  0.000
[5,]   286  2.3810       0  0.000
[6,]   292  2.0273       0  0.010
[7,]   134  2.6864       0  0.000
[8,]   320  2.2902       0  0.020

> ga.closed$significant

[1] "androgen_receptor_signaling" "apoptosis"
[3] "cell_cycle_control"

> ga.closed$not.significant

[1] "notch_delta_signalling"
```

We get the same significant and non significant pathways as before.

6 Diagnostic Plots

There are two types of diagnostic plots available supporting communication and interpretation of results of the global ANCOVA. The `Plot.genes` visualizes the influence of individual genes on the test result while the `Plot.subjects` visualizes the influence of individual samples. Both plots are based on the decomposition of sums of squares.

We use again the van t'Veer data constricted to the genes of the p53–signalling pathway for demonstration of the plot functions.

6.1 Gene Plot

The influence of each gene on the outcome of the test can be assessed and visualized with a diagnostic plot generated by our function `Plot.genes`. It corresponds to the function `geneplot` in the *globaltest* package. The function `Plot.genes` gives a graphical display of single gene-wise analysis for all genes. Bars are always positive as a reduction of sum of squares is always achieved in this case. The bar height indicates the influence of the respective gene on the test statistic. The added reference line is the residual mean square error per gene and corresponds to the expected height of the bars under the null hypothesis which says that the gene is not associated with the clinical outcome. The bars can be colored according to a variable of interest with the option *colorgroup* in order to show in which of the groups a gene has the highest expression values.

The commands for creating gene plots in the *GlobalAncova* and the *globaltest* are as follows. Note that for the former one again three alternatives for function calls are provided, see section 4 for details.

The two approaches show almost the same results (figures 1 and 2). We prefer plotting horizontal bars rather than vertical because we think it is easier to read off the bar heights this way.

```
> Plot.genes(xx = vantVeer[p53, ], group = metastases)
> gt.vV <- globaltest(X = vantVeer[p53, ], Y = metastases)
> geneplot(gt.vV)
```

In this case where only the influence of one variable is of interest (and therefore the easiest version of possible function calls is chosen), the same variable is assumed to be relevant for coloring. However one is free to specify another coloring. For example for the same plot we could ask which genes are higher expressed in samples with either positive or negative Estrogen receptor status, see figure 3.

```
> Plot.genes(xx = vantVeer[p53, ], formula.full = ~metastases,
+     formula.red = ~1, model.dat = phenodata, colorgroup = "ERstatus")
```

6.2 Subjects Plot

The function `Plot.subjects` visualizes the influence of the individual samples on the test result and corresponds to the `sampleplot` of Goeman. The function `Plot.subjects` gives information on the reduction of sum of squares per subject. Here we sum over genes. Large reduction demonstrates a good approximation of
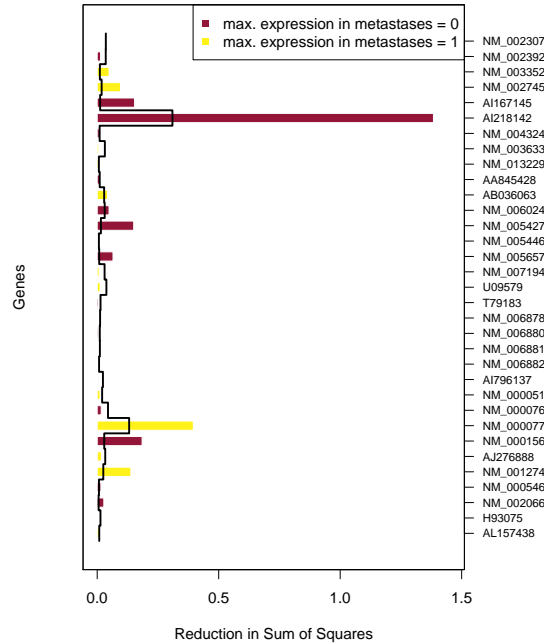
Figure 1: Gene Plot for the van t'Veer data with *GlobalAncova*. Shown are the genes of the p53–signalling pathway. The bar height indicates the influence of the respective gene on the test statistic. The colour shows in which of the phenotype groups the gene has higher expression values. The reference line is the residual mean square error per gene.

a subject's gene expressions by the corresponding group means. If an individual does not fit into the pattern of its phenotype, negative values can occur. A small p–value will therefore generally coincide with many positive bars. If there are still tall negative bars, these indicate deviating samples: removing a sample with a negative bar would result in a lower p-value. The bars are colored to distinguish samples of different clinical entities that can again be specified by the user through the option *colorgroup*. With the option *sort* it is also possible to sort the bars with respect to the phenotype groups.

We compare again the different approaches (figures 4 and 5):

```
> Plot.subjects(xx = vantVeer[p53, ], group = metastases,
+     legendpos = "bottomright")
> sampleplot(gt.vV)
```

The function `Plot.subjects` can be invoked by the three alternative function calls (see section 4) and hence also plots corresponding to more complex testing challenges can be produced as well. To give just one example we consider again the influence of the tumor grade, which can take three possible values, on gene expression (figure 6).
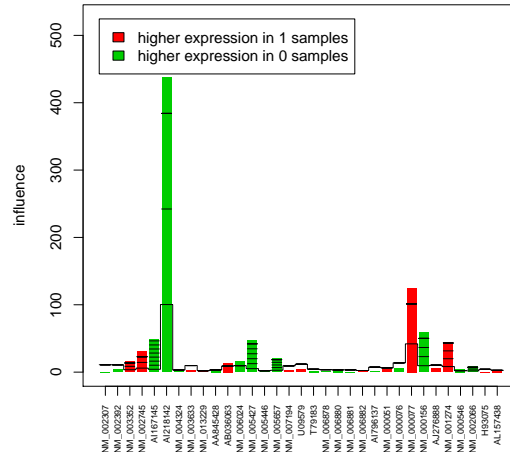
16

Figure 2: Gene Plot for the van t'Veer data with *globaltest*. Shown are the genes of the p53–signalling pathway. The bar height indicates the influence of the respective gene on the test statistic. The colour shows in which of the phenotype groups the gene has higher expression values. The reference line gives the expected height of the bar under the null hypothesis. Marks indicate with how many standard deviations the bar exceeds the reference line.

```
> Plot.subjects(xx = vantVeer[p53, ], formula.full = ~ordered(grade),
+     formula.red = ~1, model.dat = phenodata, colorgroup = "grade",
+     legendpos = "topleft")
```

7 Acknowledgements

References

[1] J. J. Goeman, F. de Kort, S. A. van de Geer, and J. C. van Houwelingen. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*, 20 (1):93–99, 2004.

[2] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, pages 531–537, 1999.

[3] S. Holm. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, 6:65–70, 1979.
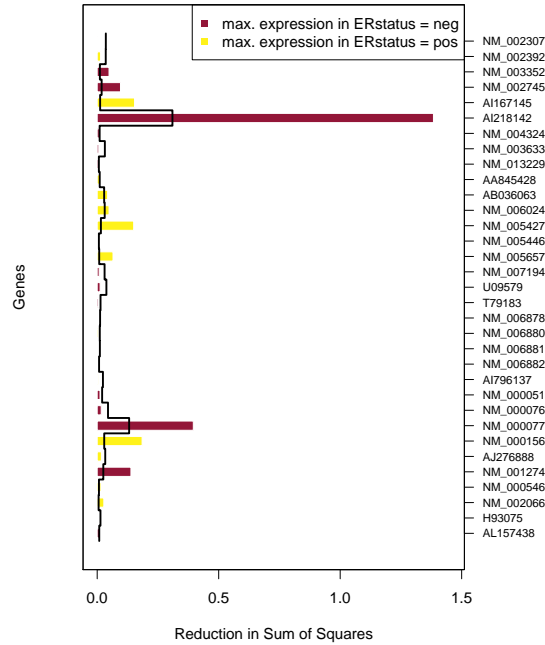
Figure 3: Gene Plot for the van t'Veer data with *GlobalAncova*. Shown are the genes of the p53–signalling pathway. The bar height indicates the influence of the respective gene on the test statistic. The color shows in which of the specified phenotype groups, in this case Estrogen receptor status, the gene has higher expression values. The reference line is the residual mean square error per gene.

[4] U. Mansmann and R. Meister. Testing differential gene expression in functional groups. *Methods Inf Med*, 44 (3), 2005.

[5] R. Marcus, E. Peritz, and K. R. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63 (3):655–660, 1976.

[6] L. J. van t'Veer, H. Dai, M.J. van de Vijver, Y.D. He, A.A.M. Hart, M. Mao, H.L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, G.J. Schreiber, R.M. Kerkhoven, C. Roberts, P.S. Linsley, R. Bernards, and S.H. Friend. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415:530–536, 2002.
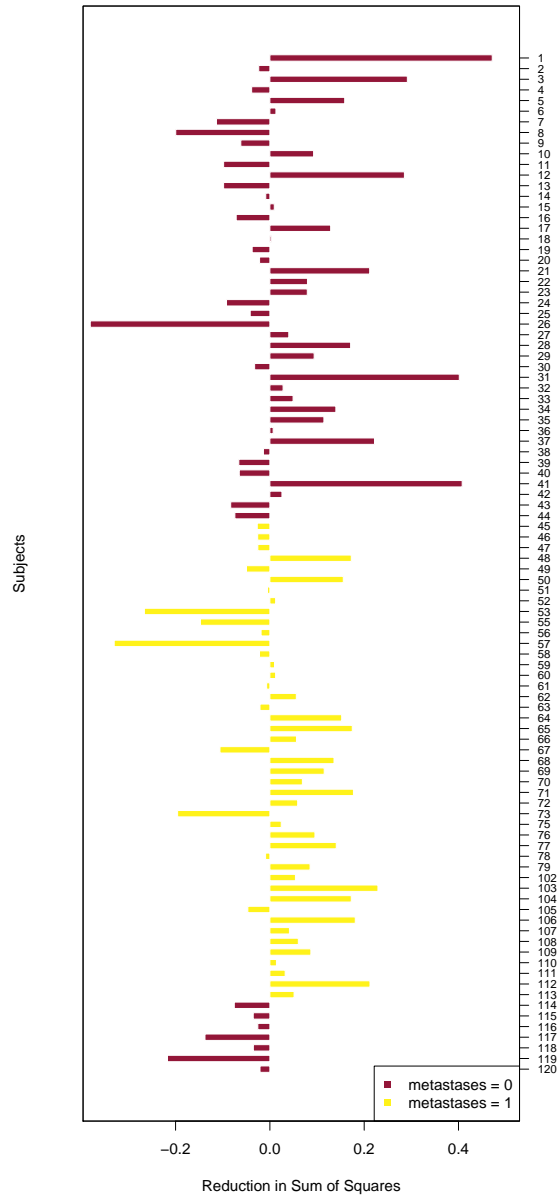
Figure 4: Subjects Plot for the van t'Veer data with *GlobalAncova*. The bar height indicates the influence of the respective sample on the test result. If an individual does not fit into the pattern of its phenotype, negative values can occur. Bars are colored corresponding to phenotype groups.
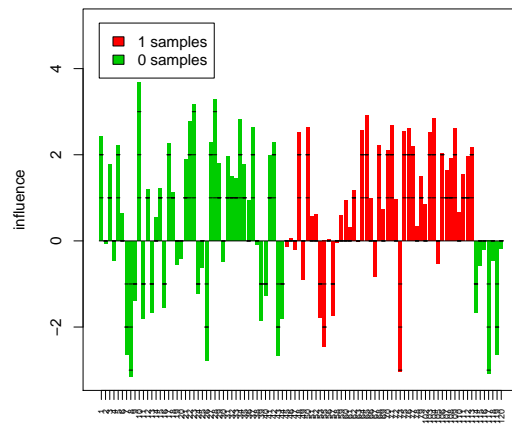
19

Figure 5: Subjects Plot for the van t'Veer data with *globaltest*. The bar height indicates the influence of the respective sample on the test result. If an individual does not fit into the pattern of its phenotype, negative values can occur. Bars are coloured corresponding to groups. The reference line shows the expected influence of the samples under the null hypothesis. Marks on the bars indicate the standard deviation of the influence of the sample under the null hypothesis.
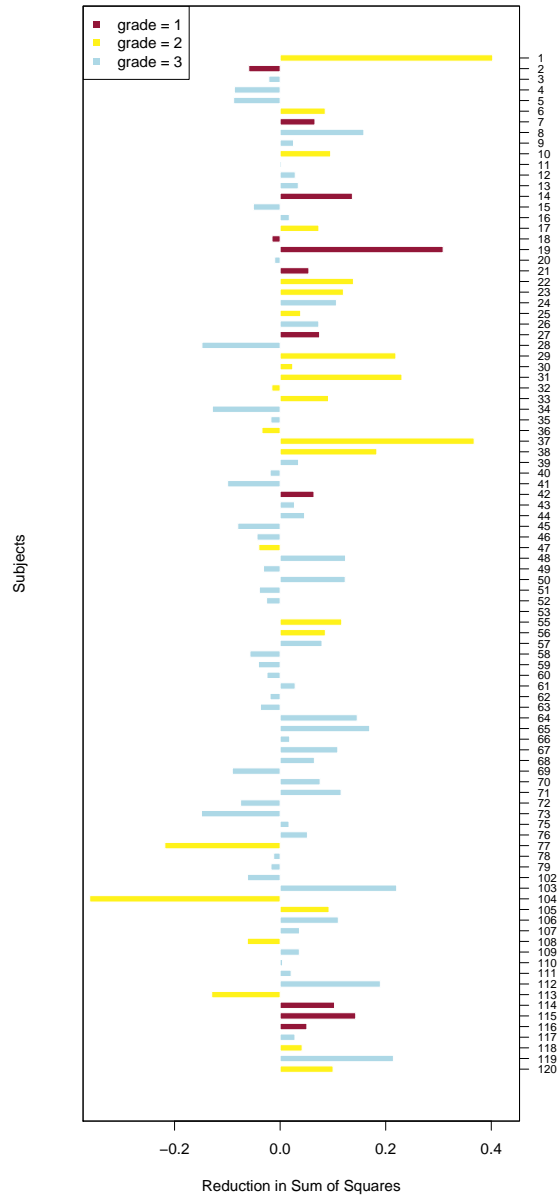
Figure 6: Subjects Plot for the van t'Veer data with *GlobalAncova*. Tumor grade is the clinical variable of interest. The bar height indicates the influence of the respective sample on the test result. If an individual does not fit into the pattern of its phenotype, negative values can occur. Bars are colored corresponding to phenotype groups.