University Medical Center Göttingen

Georg-August University

Master of Science Program

Molecular Medicine

Evaluating miRNA and target-mRNA Expression Profiles

Master Thesis

submitted by

Stephan Artmann

Göttingen 2011

General Information

Title	Evaluating miRNA
	and target-mRNA Expression Profiles
Department	Department of Medical Statistics
	University Medical Center Göttingen
Name, First Name	Artmann, Stephan
Matriculation Number	20662324
Address	Rosenbachweg 4/305, 37075 Göttingen
Phone	0176 / 6 25 25 25 7
E-Mail	stephan.artmann@stud.uni-goettingen.de
Supervisor	Tim Beißbarth
Contact	Department of Medical Statistics
	University Medical Center Göttingen
	Humboldtallee 32
	D-37099 Göttingen
Phone	
E-Mail	
Second Evaluator	
Contact	
Phone	
E-Mail	
Date of Submission	

Table of Contents

Abbreviations

Abbreviation	Denotation
APR	Average Power Rate
FDR	False Discovery Rate
FWER	Family-wise Error Rate
KS	Kolmogorov Smirnov
mRNA	messenger-RNA
miRNA	micro-RNA
nt	nucleotide(s)
RISC	RNA-induced silencing complex
UTR	untranslated region

Abstract

Interest in micro-RNAs (miRNAs) has been constantly growing in recent years. This Thesis concentrated on two-group testing in high-throughput data, i. e. on the question which miRNAs lead to or are part of differences between two groups. A straightforward approach would be to test each miRNA on its own, for example with the well-known t-test. However, if mRNA expression data is present as well, such analyses do not cover all the information available and therefore unnecessarily lose power. The intention here was to incorporate both the miRNAs' as well as their target-mRNAs' expression values. p-values derived from miRNA-wise testing as well as by testing for differential expression in the gene set lead to a statistic describing the overall probability that a miRNA or its respective gene set change in their expression between two groups. Different algorithms to achieve this were compared in simulations. The pipelines were used on real expression data and their ability to refine the analysis was shown. The algorithms were implemented in the publicly available 'miRtest' R package to allow further use.

Das Interesse an micro-RNAs (miRNAs) ist in den letzten Jahren rapide gewachsen. Diese Arbeit behandelt den Vergleich zweier Gruppen, von welchen high-throughput-Daten vorliegen, um die Frage zu klären, welche miRNAs zwischen diesen in ihren Expressionwerten differieren. Ein simples Verfahren wäre jede miRNA einzeln, etwa mit dem klassischen t-Test, zu testen. Sofern auch mRNA-Expressionsdaten zur Verfügung stehen würde dies jedoch nicht alle Informationen ausnutzen, die vorhanden ist. Somit würde beim Testen unnötigerweise Macht verschenkt. Die hier verfolgte Strategie versucht, miRNA- und mRNA-Expressionswerte zusammenzufassen und gemeinsam zu betrachten. p-Werte aus miRNA-weisen Tests sowie aus gene set tests führen zu einer Statistik, welche die Wahrscheinlichkeit, dass eine miRNA zu Effekten zwischen zwei Gruppen führt, besser beschreibt. Verschieden Algorithmen dies zu bewerkstelligen wurden anhand synthetischer Daten verglichen. Anhand eines microarray-Datensatzes wurde gezeigt, wie die Analyse durch solcherart Algorithmen verfeinert wird. Letztere werden in einem öffentlich zugänglichen Paket namens 'miRtest' zur Verfügung gestellt.

1 Introduction

1.1 Motivation

The discovery of miRNAs took place a little more than a decade ago. Since than, a rapidly growing field has emerged which concentrates on their understanding. More and more details of the pathway they trigger are being elucidated while a growing number of their functions is being identified. Even though their role in mammalian cells is not yet entirely understood, they appear to play a crucial role in development and cancer (Lu *et al.*, 2005; Volinia *et al.*, 2006). Indeed, it was shown (Lu *et al.*, 2005) that tumour samples, for example, can be clustered better when the underlying miRNA expression profile is known, than with a usual mRNA profile. Despite their importance in plant and animal cells, it was not for methodology why they were discovered so late. Instead, it was the idea that RNAs as short as miRNAs are more than just degraded mRNAs occurring in an experiment.

The analysis of miRNA expression is rather straightforward: all the methodology available for regular mRNAs just had to be adapted to allow miRNA purification. Still, it is complicated and expensive in comparison to their messenger pendants. miRNAs can be quantitatively analysed in RT-PCR. Alternatively, they are often quantified in a high-throughput manner on special microarrays which are gaining popularity at the moment. Finally, standard biostatistical methods, such as linear models in group comparison, can be applied to their expression values. This, however, does not include all the information we have on them. We know, for example, that they are part of special and simple pathways: they lead to the down-regulation of their targets rather directly. Consequently, inferring information from a-priori defined gene groups, or gene sets, that are supposed to be influenced by a miRNA appear to be worthwhile. Methods to analyse gene sets have been coming up in recent years as part of pathway analysis (see, for example Subramanian *et al.*, 2005).

Ideas to combine gene set testing with the miRNAs' own expression profile have been rare, however. See (Jayaswal *et al.*, 2009) for an example. The intention of this Thesis

is to combine all the information that is there. p-values from miRNA-wise two-group comparisons shall be combined with p-values coming from their gene sets, which can be defined from target prediction data. The idea is to provide a score or even p-value which allows easy interpretation of whether a miRNAs affects a cell's/tissue's/organism's phenotype.

1.2 miRNA

The only about 20 nt long transcripts pose a comparatively new mode of gene regulation. By down-regulation of their targets, which they recognise specifically by their sequence, they can have a large impact on a cell's gene expression pattern. In the following, miRNA expression, maturation and its function shall be summarised.

1.2.1 Biogenesis

As every RNA, pri-miRNAs, the predecessors of mature miRNAs, are transcribed in the nucleus. They can be located within introns of protein coding genes (Rodriguez *et al.*, 2004) or in intergenic region. Indeed, a primary transcript, which is capped and has a poly-A-tail, can be coding (CAI *et al.*, 2004). Their transcription is supposed to be mediated by Pol II (Lee *et al.*, 2004). This often happens in a polycistronic manner (Chang and Mendell, 2007). A characteristic of the pri-miRNA are its stem loops, 60 to 80 nt long regions which fold back to themselves. They are the ones where the later miRNAs are located. Already in the nucleus they are freed from the pri-miRNAs and the parts in between are degraded quite efficiently (Chang and Mendell, 2007). The abundance of pri-miRNA in cells is therefore usually quite low. This processing step is mediated by the microprocessor complex (Seitz and Zamore, 2006). The main effector in it is the endonuclease Drosha. The stem loops, also called pre-miRNAs, have a 3' overhang which allows them to be exported into the cytoplasm via exportin 5 (Bohnsack *et al.*, 2004). There, they encounter Dicer, an enzyme which removes the loop of the hairpin turning it into a double stranded RNA (Ambros *et al.*, 2003). This can be incorporated into the RNA induced silencing complex RISC, the function of which is described in the

next chapter. Here, one of the two strands need to be degraded to allow the remainder to convey its function. Usually, one of the two strands is degraded more frequently than the other. It is then denoted by a star (*) in its name.

1.2.2 Mode of Action

It is often said that a miRNA attacks or degrades its targeted mRNAs. However, this is only partly true, as the role of the miRNA is rather to recognise the targets while RISC is the one to lead them to their fate. Note that both are not consumed in the reaction, explaining the catalytic behaviour of miRNAs and its great impact on its target's abundance. In plant cells, where miRNAs usually show a high sequence homology to their targets RISC mainly degrades them itself. In mammalian cells, however, its mode of action appears to be more complex, or at least to be understood in greater detail. Here, RISC can also cleave mRNAs, mainly via one of the proteins it is composed of, Argonaute2 (Liu *et al.*, 2004; Meister *et al.*, 2004). However, it can also do modifications which lead to mRNA degradation later on, including decapping (Eulalio *et al.*, 2007) and deadenylation (Eulalio *et al.*, 2009; Wu *et al.*, 2006). Another mode of action is the transport of mRNAs to P-bodies, where they can be degraded or stored (Cannell *et al.*, 2008). Finally, RISC can also just repress the translation of a mRNA. However, as the the last two points do not affect total mRNA levels of a cell, they can only be measured using proteome-based technologies. On a typical microarray platform, however, these effects are not detectable.

1.2.3 miRNA Target Prediction

As mentioned before, miRNA-mRNA interaction is sequence dependent. However, due to mostly imperfect sequence homology between the two the in-silico identification of miRNA targets is more challenging than just a search for matching sequences. Different methods have been developed to allow target prediction. Examples for databases that provide such information are 'miRbase' (Griffiths-Jones *et al.*, 2006) or 'TargetScan' (Grimson *et al.*, 2007). Some features most of them have in common are:

In the 3'UTR, the 3' untranslated region, of a mRNA a so called seed sequence needs

to be located which is thought to have a great influence on the probability of a mRNA being degraded (Brennecke *et al.*, 2005). Another point for deciding in favour or against a mRNA being a target is the expected stability of the interaction with a miRNA. The longer a miRNA is able to keep a mRNA close to RISC, the higher the probability that it will interact with it. Apart from that one can infer information from different species: if, in general, sequences in a mRNA are conserved between related species one can expect that they have a crucial role and their change in function or their lacking leads to an evolutionary disadvantage. Accordingly, target sites of a mRNA that are more conserved may be expected to have a higher importance for mRNA-miRNA pairing. Finally, other points include the number of target sites that a mRNA has.

Despite all the information available from the sequence target prediction is still complicated and the results from different algorithms can be quite different (Sethupathy *et al.*, 2006). This is one reason why it has been suggested by Sethupathy *et al.* (2006) to combine different predictions and rather allow more false positives than false negatives. Another disadvantage is that such algorithms cannot know whether a certain mRNA is present in a cell at all so that it can indeed be attacked by a miRNA. Finally, target prediction algorithms just return p-values describing the probability of a miRNA attacking an mRNA, which not necessarily allows the deduction of how efficient that down-regulation is.

1.3 High-Throughput Data Analysis

1.3.1 Microarrays

The most popular high-throughput method to measure RNA abundance is the microarray technology. Most often it is a slide of single stranded DNA spotted on it. To measure RNA abundance in a biological sample, RNA is reverse transcribed into cDNA which is labelled with a fluorescent marker. The resulting targets are hybridised on the array. A scanner determines the strength of the fluorescence signal in every spot and therefore allows the relative quantification of each spot.

In the optimal case, the sequence of each spot should be complementary to one transcript, or at least, one gene only. This, however, is often not the case. There are further

methodological issues blurring microarray results (Nykter *et al.*, 2006). The binding efficiency of targets on the probes can vary. So can the size of the spots. The quality of some spots may be poor, leading to spots suffering from a chord cut or being close to one or more of their neighbours. Also, problems when scanning the microarray can occur (Nykter *et al.*, 2006). All of this as well as further issues lead to a variance between technical replicates, which adds up to the biological variance one sees when taking samples from different individuals. Consequently, the distribution of fluorescence intensity differs for every repetition of a microarray experiment, even if the same platforms are used. It has therefore become a standard procedure to normalise microarray expression data, i. e. to make the fluorescence values of all microarrays follow the same distribution. There exist several methods to achieve this. The quantile normalisation which will be used in this Thesis (see Material & Methods) normalises microarrays without changing gene ranks.

Recently, more and more companies also offer miRNA microarrays. They follow the same principle as usual mRNA-arrays but are specialised for miRNA quantification.

1.3.2 Multiple Hypothesis Testing

A *p*-value is generated by transforming a given statistic to a uniform distribution between 0 and 1. It describes the probability of a statistic being as extreme as or more extreme than observed. In microarray studies, however, where thousands of genes and hundreds of gene sets are involved, the probability of observing extreme statistics just by chance increases dramatically.

To deal with this problem one first has to define error rates. Most of the following information is based on the review by Dudoit *et al.* (2003). For a single experiment, a type I error is committed when a null hypothesis is rejected even though it is true. On the contrary, a type II error means not declaring a test significant while the alternative hypothesis is true. When one switches to more-dimensional experiments, as is the case for high-throughput measurements, one defines rates. The most stringent of these is the family-wise error rate $FWER$, i. e. the probability of having maximally one false positive detection. The most famous method for $FWER$-control is the procedure described by

Bonferroni (Bland and Altman, 1995). It is very simple yet very conservative as well. Bonferroni-adjustment works by multiplying every p-value p_k of an experiment k, here of a gene k, with the number n of genes tested. Other ideas which detect more alternative hypotheses exist, but generally $FWER$-control is unnecessarily conservative. When conducting a microarray experiment, one is usually interested in controlling the proportion of false positives among the positives, the false-discovery rate FDR (Benjamini and Hochberg, 1995). It is important to differentiate between weak and strong control of an error rate here. The latter refers to the case that a procedure controls an error rate for any combination of true and false null hypotheses, while in the first case, it is only controlled when all null hypotheses are true. A procedure that has strong control over the FDR has a weak control over the $FWER$. Different procedures have been proposed that lead from p-values to so called q-values (see below). In a list of genes ordered according to their p-values, a q-value of a given gene i describes the expected FDR if one rejects the null hypotheses of all the genes in the list from the first up to gene i.

The method of Benjamini and Hochberg (1995) is one of the most popular ones and it will also be applied here. The idea is that if one has an ordered list of p-values, one does not need to control all of them according to Bonferroni. This is sufficient for the lowest one. For every following p-value p_k, where k is its rank among the others, Benjamini and Hochberg (1995) find K as the largest k which satisfies

$$p_k \leq \frac{k}{n} \cdot \alpha,$$

where α is the threshold of significance applied and $n = \max(k)$. For the hypotheses $k \in \{1, ..., K\}$ the null is rejected. Additionally, one may be interested in the q-value, that is the smallest FDR-threshold one has to assign to a gene to call it significant:

$$q_k = \min\left(\min\left(\frac{n}{k} \cdot p_k, 1\right), q_{k+1}\right)$$

for all $k \in \{1, ..., n-1\}$ and

$$q_n = \min\left(\frac{n}{n} \cdot p_n, 1\right) = \min\left(p_n, 1\right).$$

Note that the order of genes is not changed by the recursive formula: the minimum of each gene's and its successor's q-value is taken. This way the q-value becomes a monotone

function of the *p*-value. Consequently, after the adjustment only the threshold below which genes are declared significant is shifted towards the more significant genes. It was shown by Benjamini and Yekutieli (2001) that this controls the expectation value of the FDR under the assumption of positive regression dependency. This includes data with positive covariance matrix, such as autoregressive covariance (see Section 2.6.1). A method to adjust for any dependencies between the genes was proposed by Benjamini and Yekutieli (2001) as well. It is a modification of the method above. As it has no assumption whatsoever on the correlation structure it has to assume the worst case of correlation which makes it a very conservative method.

While being in control of type I errors, one wishes to make as few errors of type II as possible, i. e. to have a large power rate. Analogous to the different type I error rates, different power rates have been defined. Three common ones are: The probability of correctly rejecting at least one null hypothesis, the average power rate APR defined as the average probability of rejecting a false null and the probability of rejecting all null hypotheses that are false. It will be the APR that will be utilised here.

1.3.3 Important Distributions of Random Variables

In the following chapters different tests will be presented and used. To be able to generate *p*-values they need to arrive at a statistic of known distribution. If assumptions on the distribution of the original data are made (it is most often required to follow a normal distribution then) a test is called parametric. Else, it is a non-parametric test. Most of the following was gathered from Wei (2010). Whenever other sources were used they are indicated.

The most famous distribution is the normal distribution Wei (2010). According to the central limit theorem the mean of random variables, almost independently of their distribution, asymptotically (for more or less large numbers of repetitions) follows a normal distribution. In the univariate case, its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2 \cdot (\frac{x-\mu}{\sigma})},$$

where μ and σ are the two parameters of the function, the expectation and the variance, respectively. Two or more normally distributed variables can, but not necessarily do, follow a multivariate normal distribution. It is the generalisation of the normal distribution to higher dimensions. Its expectation vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ parametrise the density function:

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \cdot e^{-1/2 \cdot (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}.$$

The covariance matrix has the variance of each element of $\boldsymbol{x}$ on its main diagonal. The covariance of the ith and the jth element of $\boldsymbol{x}$ is in elements $(\sigma_{i,j})$ and $(\sigma_{j,i})$ of $\boldsymbol{\Sigma}$. If the corresponding entry is 0, it means not only that the two variables are linearly uncorrelated, but that they are indeed independent. $\boldsymbol{\Sigma}$ is a symmetric, positive definite matrix. The covariance $\sigma_{x,y}$ of two variables x and y is proportional to their correlation $\rho_{x,y}$ according to $\rho_{x,y} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y}$.

In its simplest form, a χ^2-distribution describes the distribution of the square of a standard normally distributed variable. However, also a sum of squares of independent standard normally distributed variables follow a χ^2-distribution. Its only parameter is the number k of variables the square sum of which was summed up. The density function is

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} \cdot x^{k/2-1}e^{-x/2} \text{ if } x \geq 0,$$

$$f(x;k) = 0 \text{ otherwise,}$$

where Γ is the gamma distribution function. For a multivariate distribution the pendant to the χ^2-distribution is the Wishart-Distribution.

Some further distributions are necessary to entirely understand Section 2: One is the Bernoulli distribution which can take either value 0, or value 1, the latter with a certain probability p. It will be used in section 2 to randomly sample logical vectors. The distribution of the number of Bernoulli trials necessary to get one success (one value 1) follows a geometric distribution. It is the probability that the kth out of k trials is the first success

$$Pr(X = k) = (1 - p)^{1-k}p,$$

where X is the number of trials needed to get one success. The hypergeometric distribution is somewhat similarly motivated. It describes the probability that in an experiment without replacement, the number of successes X is k for a given number n of trials out of N possible trials:

$$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}.$$

Finally, the t-distribution is of interest here. It occurs for a variable Z if a normally distributed variable is divided by a variable following a χ^2-distribution. The distribution often occurs in two-group testing. The mean $\overline{X}$ of independent normally distributed samples follows a normal distribution if X does. For a known variance σ one can therefore construct the statistic

$$S = \frac{\overline{X}_1 - \overline{X}_2}{\sigma} \cdot \sqrt{(n)},$$

where X_1 and X_2 are the mean of the first and the second group, respectively. S then follows a standard normal distribution. However, as σ is often unknown it has to be replaced by an estimator s which is generated from the data and follows a χ^2-distribution. In the case of equal variances in both groups the test statistic then becomes

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{s} \cdot \sqrt{(n)}$$

and follows a t distribution.

1.3.4 Gene Set Testing

Tests for differential expression have classically been applied in a gene-wise manner. However, lists of genes ordered according to the significance against the null hypothesis are usually not very informative. Consequently, approaches that calculate statistics for pre-specified sets of genes have been developed in recent years. They have been grouped into self-contained and competitive tests by Goeman and Bühlmann (2007). The first build a statistic based solely on the information present inside a given gene set. One of the first to do so was the 'globaltest' by Goeman *et al.* (2004). They test against the null hypothesis that no single gene in the set is differential. Competitive tests, on the contrary, compare the differential expression of the genes inside the gene set with those outside of

it. They help to make a decision whether the expression distribution of the gene set's genes is the same as of those outside the gene set. They are therefore more robust to sporadic effects which are not restricted to the gene set, e. g. a change in expression that occurs in all genes. For a given experiment, their p-values for the different miRNAs can be expected to follow a somewhat two-tailed distribution under the alternative: the genes of those gene sets which are significant lead to higher p-values of the gene sets that do not. Nevertheless, for the false null hypotheses one might expect them to have lower power than self-contained tests.

2 Material and Methods

2.1 miRNA-wise Testing using Linear Models

To calculate miRNA-wise *p*-values the popular linear models by Smyth (2004) were applied. They are implemented as a part of the 'limma' package, which is written in the R statistical programming language (R Development Core Team, 2010). In the following, the notation of Smyth (2004) will be a little adapted to explain their model.

Define a response vector $y = (y_i)$, where $i = 1, .., n$ is the ith microarray replicate. Here, it will be a vector of two factors $k = 1, 2$ indicating which group microarray i belongs to. Let the matrix X of gene expression values consist of g row vectors α_j, where $j = 1, ...g$ is the index corresponding to a certain gene. It is assumed that $E(y) = D \cdot \alpha$, where D is a so called design matrix. Here, an element $d_{i,k}$ of D will be 1 if microarray i is in group k, and 0 otherwise. One is interested in the coefficients β_j of the linear model $\beta_j = C \cdot \alpha$, where C is a contrast matrix. Here, it is simply designed in such a way that each β_j reflects the difference between the two groups. Assuming a normal distribution of β_j and a χ_2 distribution of its variance v_j a statistic t_j is calculated by

$$t_j = \frac{\hat{\beta}_j}{s_j \cdot \sqrt{v_j}},$$

where s is the estimator of a gene's variance. By assuming that the estimator of every gene's coefficient follows a normal and the estimator of a gene's variance a scaled χ^2-distribution,

$$\hat{\beta}_j \sim N(\beta_j, v_j s_j^2) \text{ and } s_j^2 \sim \frac{\sigma_j^2}{d_j} \cdot \chi_{d_j}^2,$$

t_j follows an approximate *t*-distribution with d_j degrees of freedom. Smyth (2004) further takes into account that the analysis of genes happens in a very parallel manner. This is done by application of a hierarchical model. Essentially, prior distributions of the unknown coefficients β_j and σ_j^2 across genes are assumed:

$$\frac{1}{\sigma_j^2} \sim \frac{1}{d_0 s_0^2} \cdot \chi_{d_0}^2 \text{ and } \beta_j | \beta_j \neq 0 \sim N(0, v_{0j} \sigma_j^2).$$

Under the hierarchical model, a so called moderated $\tilde{t}$-statistic is defined where s_j is replaced by $\tilde{s}_j$,

$$\tilde{t}_j = \frac{\hat{\beta}}{\tilde{s}_j \sqrt{v_j}}, \ \tilde{s}_j^2 = \frac{d_0 s_0^2 + d_j s_j^2}{d_0 + d_j}.$$

Clearly, values for the hyperparameters s_0 and d_0 are needed. Smyth (2004) estimates them from the data. Under the null hypothesis that $\beta_j = 0$, $\tilde{t}_j$ follows a t-distribution with $d_j + d_0$ degrees of freedom.

2.2 Gene Set Tests

2.2.1 Global Tests

The three R packages applied here were 'globaltest' (Goeman *et al.*, 2004), 'GlobalAncova' (Mansmann and Meister, 2005; Hummel *et al.*, 2008) and a test based on repeated measures named 'RepeatedHighDim' (Brunner, 2009). They all implement what may be called self-contained testing (Goeman and Bühlmann, 2007), i. e. that only data from inside the gene set is used for the test. The corresponding null hypothesis can be formulated as that no single gene is differentially expressed in a gene set. All of them lose information on directionality of the gene set's differential expression when calculating their test statistic. For miRNA testing, however, it is of interest to include this information as one knows that miRNAs and their target mRNAs should act inversely when it comes to differential expression. To avoid this drawback an approach of splitting the gene sets into up- and down-regulated genes was followed.

'globaltest' implements a procedure based on the generalised linear model of McCullagh and Nelder (1989). The authors look for a predictive effect of gene expression Y on the group membership G. Their null can be formulated as $P(G|Y) = P(G)$. To test it Goeman *et al.* (2004) ask whether the regression coefficients of their linear model are zero. After assuming that the regression coefficients are sampled from a common distribution with expectation zero, they reformulate their null as to whether the variance of that distribution is zero. For their test they arrive at a statistic Q approximately following a scaled χ^2-distribution, which can be reformulated as the mean of the squared covariances

between each gene's expression and the membership variable (the latter is a vector that describes which sample belongs to which of the two groups). Consequently Q can lead to two-sided p-values only.

While 'globaltest' tests how gene expression affects the group membership, (Mansmann and Meister, 2005; Hummel *et al.*, 2008) look at the role G has on gene expression. Their null hypothesis may be formulated as $P(Y|G = 1) = P(Y|G = 2)$. This is equivalent to the one of Goeman *et al.* (2004), however, under the global null hypothesis. The test is also based on the analysis of covariances. An observed expression value x is modelled as being composed of a mean μ, a per-sample covariate z and an error e with expectation $E(e) = 0$. The mean μ itself is split into a group effect α, a gene effect β and their interaction, γ: $\mu = \alpha + \beta + \gamma$. The null hypothesis tested is that $\gamma = 0$. An F statistic is derived reflecting this null which can be transformed into a p-value. Note, however, that the test assumes independence of genes. To cope with that subject permutation was proposed by Mansmann and Meister (2005). However, due to small sample sizes, the approximative method of p-value calculation was used here.

The third approach of global testing is a general model for high-dimensional repeated measures data (Brunner, 2009), in the following referred to as RepeatedHighDim or RHD. This approach tests the same hypothesis as GlobalAncova.

2.2.2   Enrichment Tests

Enrichment tests work on what is regarded as a competitive null hypothesis by Goeman and Bühlmann (2007). Originally, a running sum statistic was proposed, which turned out to be a weighted Kolmogorov Smirnov (KS) statistic , under the name gene set enrichment analysis (GSEA, Subramanian *et al.*, 2005). The Kolmogorov Smirnov statistic, however, had originally been designed for independent replicates of an experiment. As this assumption usually does not hold for genes a resampling procedure was proposed (Shi, 2007), permuting the group label of a microarray experiment. This procedure is limited by the sample size, though. A typical microarray experiment has three to four replicates in each group. The lowest p-value which one could then show would be around 5% to

1% (assuming equal sample sizes in both groups). Consequently, non-resampling-based procedures were followed in this Thesis. For all genes gene-wise testing was applied as implemented in 'limma' (see also 2.1). Genes were ranked according to their p-values. Subsequently, three different non-parametric tests were applied. They are all based on the assumption that inter-gene dependence is not negligible, a problem that is discussed in Gatti *et al.* (2010), for example.

The Kolmogorov Smirnov test has the null hypothesis that a given sample comes from a specified distribution . It compares the empirical distribution function F_n with a pre-specified distribution F. The test is based on the statistic D_n for which

$$D_n = \sup_x |F_n(x) - F(x)| .$$

It follows a Kolmogorov distribution. To calculate D_n, the empirical distribution function $S(x_i)$ is taken for every observation x_i. The maximal difference $|S(x_i) - F_0(x_i)|$ is calculated. Its distribution is then supposed to follow a Kolmogorov distribution. Here, it was the ranks of the genes inside the set that were tested against a uniform distribution between rank one and the maximum rank. The test was applied in a one-sided manner to reject the null only if the gene set tends to lower ranks.

The second test applied was a Mann-Whitney test. It is equivalent to the Wilcoxon two-sample test and will therefore be abbreviated as 'Wilcoxon test' from now on. Given the ranks of the genes inside and outside the gene set, it rejects the null if the two distributions are different (Wei, 2010). Given a gene set, the ranks of the genes inside the set are summarised to R_{gs} and the ranks of the remaining genes to R_{rest}. By subtracting a value dependent on the number of genes n_{gs} inside the set and the number n_{rest} outside of it a statistic U, called Mann-Whitney's U, is generated:

$$U_{gs} = R_{gs} - \frac{n_{gs} \cdot (n_{gs} + 1)}{2}, \quad U_{rest} = R_{rest} - \frac{n_{rest} \cdot (n_{rest} + 1)}{2},$$

$$U = \min(U_{gs}, U_{rest}).$$

For large enough ns U is approximately normally distributed with expectation $n_{gs} + n_{rest}/2$. Again, it was applied in a one-sided manner to detect only gene sets deviating towards lower ranks.

Finally, Fisher's exact test was used. In its simplest from it is a so called 2x2-table method, meaning that the association of two factors is assessed. Here, these are the factor indicating whether a gene is inside a gene set or not and the factor which describes the significance of the p-value for the gene. For every gene set one can therefore construct the following table:

Gene Set	Rest	
A	C	Significant
B	D	Not Significant

Again, a statistic of known distribution is then calculated:

$$\phi = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{(A+B+C+D)!A!B!C!D!},$$

which follows a hypergeometric distribution under the null hypothesis (Wei, 2010). Here as well a one-sided testing approach was applied.

2.2.3 Rotation Tests

Permutation of the group membership vector or of the expression matrix accordingly is a well-known procedure to generate exact p-values. If a test results in a statistic the distribution of which is unknown, it can be estimated via resampling. Such a statistic can be a simple gene-wise one, but also a statistic describing the differential regulation of an entire gene set.

The same applies to rotations, which are a generalisation of permutations for multivariate normal distributions. Consider the matrix representation of a gene expression experiment with n samples in its columns and g genes in its rows. Permuting the columns leaves the underlying gene distribution intact, yet generates 'random' data. A large number of such permutations consequently leads to an estimate of the null distribution of a given statistic. If we assume multivariate normal distribution of the independent microarray replicates it is parametrised by its mean vector $\boldsymbol{\mu}$ of gene expression and its covariance matrix $\boldsymbol{\Sigma}$. Using a rotation method, this structure remains constant yet again leads to random expression data. Indeed, resampling functions represent a subset of rotations.

The advantage of the latter is that their number is infinite, allowing one to show p-values as small as one wants (and to invest as much time as one wants to compute them). In the end, exact p-values p_{exact} can be calculated by counting how many p-values $p*$, that occurred during the rotations, were smaller than the original p-value p_{orig}:

$$p_{exact} = \frac{I(p* \leq p_{orig}) + 1}{n_{rot} + 1},$$

where I is the indicator function which returns 1 if the condition in its parentheses are true, and 0 otherwise, and n_{rot} is the number of rotations applied.

The two methods applied here were both part of the 'limma' R package that was also used for gene-wise testing (see section 2.1). The first is named 'ROAST' (Wu *et al.*, 2010). It is a self contained test which calculates the moderated t-statistic explained in section 2.1 for every miRNA inside the gene set. Different statistics can then be taken to reflect the global effect. Here, the mean of the gene-wise statistics was chosen. 'Romer', on the other hand, is a competitive test. It also computes a summary statistic for the gene set. However, it is then compared with summary statistics of genes outside the gene set. Again, the mean of the gene-wise statistics was chosen as the summary statistic.

2.3 *p*-value Combination

p-values are statistics that, under a given null hypothesis, follow an equal distribution between 0 and 1. Based on this property different methods have been designed to calculate a statistic summarising two or more p-values. Two ideas were followed in this Thesis and shall be described in the following paragraphs.

Fisher's method on p-value combination Fisher (1950) takes one-sided p-values and turns them into a statistic following a χ^2-distribution with two degrees of freedom. Here it was used once to create a p-value $p_{combined}^{up}$ for miRNA up- and gene set down-regulation and $p_{combined}^{down}$ for miRNA down- and gene set up-regulation

$$p_{combined}^{up} = F_{\chi^2} \left(-2 \cdot \left(\ln(p_{micro}^{up}) + \ln(p_{mess}^{down}) \right) \right),$$

$$p_{combined}^{down} = F_{\chi^2} \left(-2 \cdot \left(\ln(p_{micro}^{down}) + \ln(p_{mess}^{up}) \right) \right),$$

where F_{χ^2} is the cumulative distribution function of a χ^2 distribution with four degrees of freedom and p_{micro} and p_{mess} are the p-values of a miRNA and its gene set, respectively. To get a final p-value $p_{combined}$, the smaller of the two was taken to reflect the overall miRNA effect:

$$p_{combined} = \min(2 \cdot \min(p_{combined}^{up}, p_{combined}^{down}), 1).$$

Note that $\min(p_{combined}^{up}, p_{combined}^{down})$ involves multiple tests and therefore may be skewed even under the null hypothesis. Thus, the p-values were multiplied by two, which corresponds to p-value correction according to Bonferroni (Bland and Altman, 1995). This allows strong control over the $FWER$. Since the $FWER$ is the probability to commit at least one type I error, i. e. to call at least one of the up- or down-regulated gene set significant even though it is not, the minimum of these two adjusted p-values is properly corrected as well. $p_{combined}$ might only be skewed towards larger values. Consequently, one might expect this correction to keep the nominal level but to lead to light losses of power. Therefore, another way to combine p-values was applied to the rotation tests.

For these, one can make use of their property returning one-sided p-values which approximately sum up to 1 for larger numbers of rotations. As the miRNA-wise p-values returned from 'limma' do the same, one can apply a combining method based on Stouffer's method (Stouffer *et al.*, 1949; Marot and Mayer, 2009) to generate

$$p_{combined}^{r,up} = 2 \cdot \left(1 - \left| \Phi \left(\frac{\Phi^{-1}(p_{micro}^{up}) + \Phi^{-1}(p_{mess}^{down})}{\sqrt{2}} \right) \right| \right)$$

and

$$p_{combined}^{r,down} = 2 \cdot \left(1 - \left| \Phi \left(\frac{\Phi^{-1}(p_{micro}^{down}) + \Phi^{-1}(p_{mess}^{up})}{\sqrt{2}} \right) \right| \right).$$

Note that here Φ is the cumulative distribution function of the standard normal distribution. For tests whose one-sided p-values exactly sum up to one, the overall two-sided p-value is simply $p_{combined}^{r,up}$ (or $p_{combined}^{r,down}$, as they are equal) without the need for any further adjustment. To be completely rigorous, however, one needs to take into account that rotation p-values, because they are generated by resampling, do not follow a real continuous distribution but are always a little too large. When they become 1, Stouffer's method would lead to a combined p-value of zero. However, due to the way they are

calculated (see Section 2.2.3) they cannot become zero. Consequently, if one one-sided rotation p-value is 1, the other reaches the minimal p-value that can be shown given the number of rotations. One can assure that always the latter is used by calculating the overall two-sided p-value as

$$p^r_{combined} = \max \left({}^{r,up}_{combined}, p^{r,down}_{combined} \right).$$

Finally, in both approaches the combined p-values for all miRNAs were adjusted for multiple hypothesis testing using the method from (Benjamini and Hochberg, 1995, see also section 1.3.2).

2.4 Target Prediction Data

The 'microCosm' online database (Griffiths-Jones *et al.*, 2008) provides tables for different species of miRNA-mRNA pairs. To each a p-value is assigned reflecting the probability that the miRNA binds to the mRNA. Parameters of the simulation study (see section 2.6) where chosen to more or less resemble the number of gene sets, distribution of gene set size, gene set overlaps, etc. in the database.

'microCosm' is based on miRNA sequences from the 'miRbase' sequence database (Griffiths-Jones *et al.*, 2006) and the 'miRanda' algorithm (Enright *et al.*, 2003) for target site prediction. Briefly, 'miRanda' takes UTRs of different vertebrate species and scans them for potential target sites. This is, as already mentioned in the introduction, primarily based on sequence homology: The first is sequence complementarity in the 3'-UTRs in vertebrate genomes. Additionally, all protein-coding regions are scanned for high-scoring miRNA target sites as well. Further factors that are considered to generate a final score are the binding energy of the miRNA-mRNA-complex and evolutionary conservation of probable target sites between species. The cut-off parameters are chosen to reflect current knowledge on experimentally validated target sites concerning target site distributions, position of the pairing regions and others.

As the authors of the data example analysed (Nielsen *et al.*, 2009) based their analysis on the 'TargetScan' database (Grimson *et al.*, 2007), 'TargetScan v. 4.1' was also used for

the re-analysis. 'TargetScan' is also based on the 'miRbase' database. The analysis starts with 'miRbase' entries which are modified and to which sequences are added. The target prediction algorithm follows a similar idea as 'miRanda', namely to include sequence homology information and conservation across genomes of different species.

2.5 Algorithms

This short chapter intends to summarise the algorithms that were applied in the end. For an overview of them see the flowchart in Fig. 1.
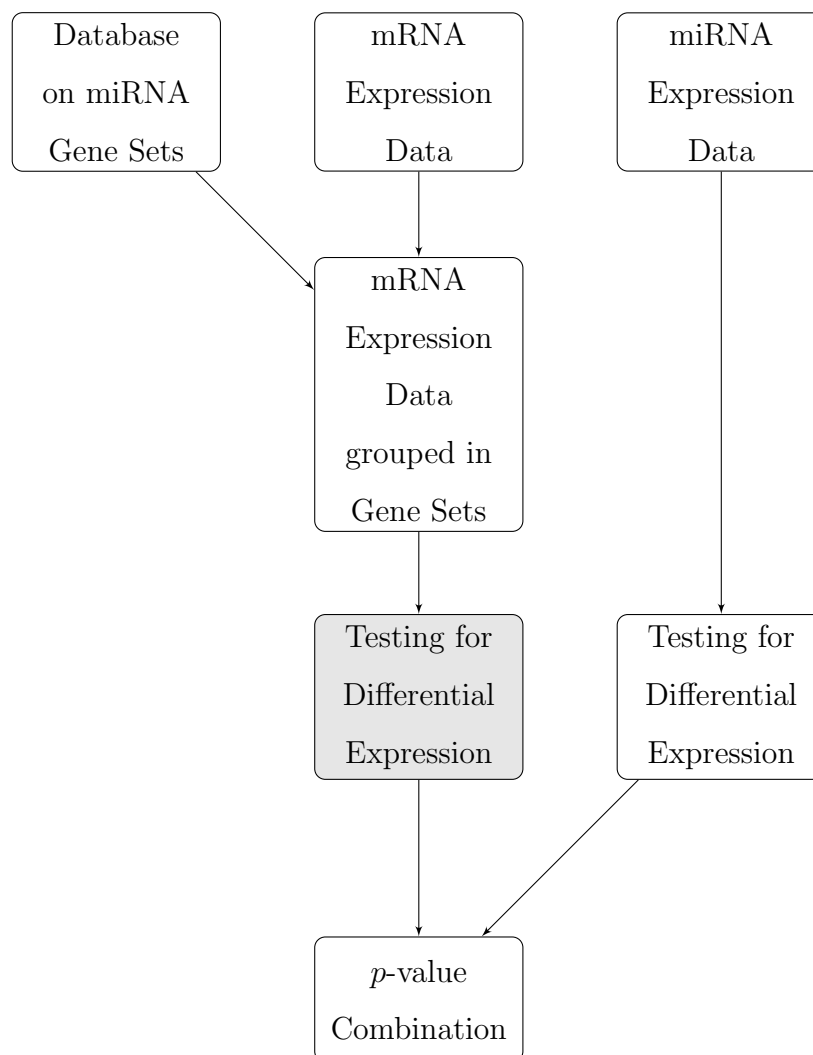

Figure 1: Flow chart of the combined testing approach. Different methods for testing for differential expression of gene sets (grey) were applied. See text for details.

The input for all of them are two miRNA expression matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ with expression values of groups 1 and 2, respectively. They have m_1 and m_2 independent replicates in their columns and d_{miR} miRNAs in their rows. Additionally, define by $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$ two corresponding mRNA expression matrices, with n_1, n_2 replicates in their columns and the d_{mess} mRNAs in their rows. Finally, the algorithm needs information on which miRNA targets which mRNAs. In the following, this shall be coded in an allocation matrix $\boldsymbol{A}$ of d_{miR} columns representing the miRNAs and d_{mess} rows for the mRNAs. Every element a_{ij} of $\boldsymbol{A}$, $i \in \{1, ..., d_{miR}\}$ and $j \in \{1, , , , , d_{mess}\}$, takes the value 1 if miRNA i leads to the down-regulation of mRNA j, and 0 otherwise.

For $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ the miRNA-wise testing described in chapter 2.1 was applied in all pipelines. The pipelines that were compared only differed in the way gene set testing was done. The gene set tests applied were the ones described above: the Global Tests implemented in the R packages 'globaltest', 'GlobalAncova' and 'RepeatedHighDim', the enrichment procedures with the non-parametric Kolmogorov Smirnov, two-sample Wilcoxon and Fisher's exact test, as well as the rotation tests 'ROAST' and 'romer'. In each pipeline the corresponding gene set test returned a list of d_{miR} p-values.

Finally, in each procedure the p-value of gene set i was combined with the corresponding ith miRNA. For all but the rotation test-based procedures Fisher's method for p-value combination (see section 2.3) was applied. For the rotation-based tests Stouffer's method was used. As quite many miRNAs were tested simultaneously, the p-values generated in this manner were adjusted for multiple hypothesis testing according to Benjamini and Hochberg (1995).

2.6   Microarray Simulation

It is impossible to decide on which pipeline works best without knowing the true null and alternative hypotheses. For a given experiment one has to know which miRNAs truly have an effect between two groups. Unfortunately, such knowledge is usually not present when dealing with real experiments. In such cases synthetic data which has effects modelled into it is often used. Such a simulation study was also applied on the pipelines presented above.

Different data structures were created. The first simulation was designed to be rather simple, evaluating the general performance of the tests. In the second, the assumption of non-overlapping gene sets was dropped to a certain degree to see how robust the methods are to that. The performance of the tests regarding the false discovery rate FDR and the average power rate APR (see Section 1.3.2 for their definitions) was evaluated. Details on how the synthetic data was generated shall be given in the following chapters. For an overview of parameters in all the simulation studies applied see Table 1.

2.6.1   Non-overlapping gene sets

The first simulation round intended to investigate the procedures' performance on rather simple data. The expression matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of miRNAs in group 1 and group 2, respectively, were drawn from a multivariate normal distribution, $\boldsymbol{X}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\boldsymbol{X}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$. The mean vector $\boldsymbol{\mu}_1$ was chosen to follow a log-normal distribution with an offset, $\boldsymbol{\mu}_1 \sim logN(0, 0.1) + 5$. In order to model an effect, i. e. a difference in means, of miRNA expression into the data, $\boldsymbol{\mu}_2$ was equal to $\boldsymbol{\mu}_1$ with the exception of a certain proportion of differentially expressed miRNAs. To one half of these, the effect strength δ was added, from the other it was subtracted. The covariance matrix in both groups was equal. $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ had four columns, representing four independent replicates in group 1 and four in group 2. The number of rows, d_{miR}, was 100, corresponding to 100 miRNAs simulated.

Next, the expression matrices $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ of mRNAs in group 1 and group 2 were created. Again, both followed a multivariate normal distribution, $\boldsymbol{Y}_1 \sim N(\boldsymbol{\nu}_1, \boldsymbol{\Psi})$ and $\boldsymbol{Y}_2 \sim N(\boldsymbol{\nu}_2, \boldsymbol{\Psi})$. To generate the mean vectors, an uninfluenced raw vector $\boldsymbol{\nu}_1^{orig} \sim logN(1, 0.1)$ was created. As a kind of noise, miRNA-independent group effects were modelled into $\boldsymbol{\nu}_2^{orig}$. Therefore, although most entries in $\boldsymbol{\nu}_2^{orig}$ were equal to those of $\boldsymbol{\nu}_1^{orig}$, a proportion of randomly sampled differential mRNAs differed by δ_{mess}. Again, to half of them this value was added while from the other it was subtracted.

To each miRNA one gene set of 50 mRNAs, which did not overlap with other gene sets, was assigned via the allocation matrix $\boldsymbol{A}$. To model the influence of $\boldsymbol{\mu}_i$ on the

corresponding $\boldsymbol{\nu}_k$, $k \in \{1, 2\}$, a linear effect was used:

$$\nu_{k,j} = \nu_{k,j}^{orig} + \sum_{i=1}^{d_{miR}} a_{j,i} \cdot \beta_{j,i} \cdot \mu_{k,i},$$

where $\beta_{j,i}$ denotes the strength (slope) of the influence of the ith miRNA on the jth mRNA in group k and $a_{j,i}$ is the corresponding entry of $\boldsymbol{A}$. $\beta_{j,i}$ was drawn from a normal distribution with negative mean m. In one simulation a weak effect was modelled with $m = -1$, in a second m was set to -10 to see whether the combination tests can control the FDR even under such strong effects. Note that since the it was the mean vectors $\boldsymbol{\mu}_k$ of miRNAs influenced the mean vectors $\boldsymbol{\nu}_k$ of mRNAs, and only then the expression matrices were drawn from the multivariate normal distribution, miRNA and mRNA replicates were modelled to be independent.

Three different types of covariance matrices were compared. In each simulation run, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ were of the same structure. The first was an autoregressive structure , meaning that the further apart two genes were, the smaller was their covariance. The main diagonal of $\boldsymbol{A}$ was set to a value constantly increasing from 1 for the first mRNA to 2 for the last. The secondary diagonals had values of 0.7 which were then constantly decreasing for every further diagonal. The second covariance structure was a block structure. This is very similar to the autoregressive structure, only that blocks in the covariance matrix behave as entries do: the farther a block of a size of 5 is away from the main diagonal, the smaller are its entries (which are all equal). The main diagonal itself was again chosen to vary from 1 to 2. Finally, an unstructured matrix was applied. The main diagonal was the same as before, but apart from that the other entries were random, so long as the matrix' positive definite character remained.

In each simulation run, the differentially expressed miRNAs and mRNAs were randomly sampled. In the simulations with non-overlapping gene sets effects caused by the position of miRNAs or mRNAs in $\boldsymbol{A}$ were avoided by randomly resampling first its columns, then its rows, thus leaving its structure intact.

Table 1: Simulation parameters. Four simulation studies were conducted: Two with overlapping Gene Sets (GS), two with non-overlapping Gene Sets. Each of the two had either a strong or a weak effect in the gene sets.

Parameter	Non-Overlapping GS		Overlapping GS	
	Strong	**Weak**	**Strong**	**Weak**
Number of Simulation Runs	1000	1000	1000	1000
Number of miRNAs	100	100	100	100
Number of mRNAs	5000	5000	5000	5000
Replicates per Group	4	4	4	4
Size of Gene Sets	50	50	variable	variable
miRNA Effect Strength δ	$\{0, 1, 2, 3, 4, 6\}$		$\{0, 1, 2, 3, 4, 6\}$	
mRNA δ_{mess}	1	1	1	1
Portion of diff. miRNAs	10 %	10 %	10 %	10 %
Portion of diff. mRNAs	5 %	5 %	5 %	5 %
Expectation vectors	$\sim \log N(1,0.1)$		$\sim \log N(1,0.1)$	
Cov. structures of	autoregressive		autoregressive	
miRNAs and mRNAs	block		block	
	unstructured		unstructured	
Mean m of miRNA Effect Strength $\beta \sim N(m, 0.1)$	-1	-10	-1	-10

2.6.2 Overlapping gene sets

In the second simulation round the intention was to drop the assumption that gene sets do not overlap, which indeed is a very unrealistic one. This is not what the gene set tests had been designed, however, so that only robustness against such occurrences could be regarded.

The entries in $\boldsymbol{A}$ were drawn in each run from a Bernoulli distribution so that gene set sizes varied around 100 to 120 genes per gene set and so that for a given gene set around 10 % of genes were shared with another given gene set. These parameters were chosen as they resembled the miRNA-mRNA network available on 'microCosm'. The other parameters were as described for the simulation with non-overlapping gene sets. Here as well, m was once set to -1 and in a second simulation to -10.

2.7 Data Example: Rat Neurogenesis

To see how the tests performed on biological data, an examplatory data set was retrieved from the public microarray repositories 'GEO' (Gene Expression Omnibus, Edgar *et al.*, 2002) and 'ArrayExpress' (Brazma *et al.*, 2003).

In their analysis, Nielsen *et al.* (2009) had purified rat neuronal progenitors and did miRNA and mRNA microarrays. Here, their data on cells from embryonic day 11 (E11) shall be compared to that of day 13 (E13). Nielsen *et al.* (2009) had done three replicates in each group for miRNAs and four for mRNAs. One of their intentions was to identify miRNAs differentially regulated between these two stages. Additionally, they regarded gene sets which they had identified using the 'TargetScan' (Grimson *et al.*, 2007) database, version 4.0. They then separated the gene sets into up- and down-regulated mRNAs and did a Fisher-test on them to seek out subsets of genes more differentially regulated than expected by chance.

As 'TargetScan' v. 4.0 was not available any longer, 'TargetScan' v. 4.1 was used to create the allocation matrix $\boldsymbol{A}$. Note that some discrepancies might therefore occur between the miRNA-mRNA network in the original publication and in the re-analysis. Expression data for miRNAs was retrieved from 'ArrayExpress' and for mRNAs from 'GEO'. Data was subsequently normalised using the quantile method (Bolstad *et al.*, 2003) and the algorithms detailed above were applied.

3   Results

3.1   Simulation Results

In the following the results for the simulation shall be summarised. Where necessary, figures are given for illustration. Due to the large number of simulations conducted, however, the reader shall be referred to the Appendix for a complete list of figures. Additionally, the main results and most important findings are summarised in Section 3.1.5 as well as in Tab. 2 and Tab. 3.

3.1.1   Non-Overlapping Gene Sets and weak Gene Set Effect

This simulation round was very idealised. All gene sets had the same size and did not overlap. At first the results for the autoregressive covariance structure shall be regarded.

As far as FDR control is concerned, the miRNA-wise approach always stayed considerably close to the level of 5 %. This result was confirmed under all circumstances and in all simulations conducted and will therefore not be mentioned in the following.
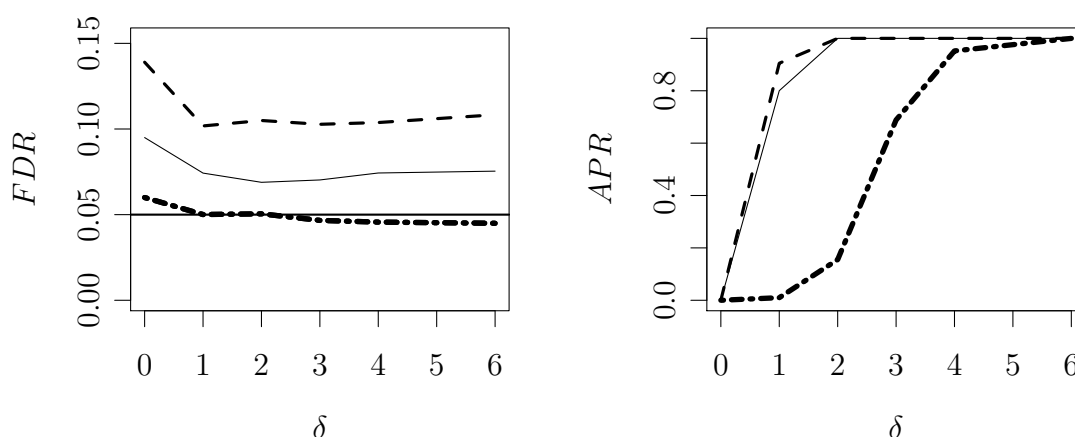

Figure 2: Simulation performance of 'globaltest'. Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

An example for this can be seen in Fig. 2, where the performance of the 'globaltest' and the combination procedure based on it is depicted. While the first is a little liberal, the latter at least compensates for this behaviour and is closer to the 5 % threshold. Still, it is not able to entirely control the FDR. As far as the APR is concerned, it is greatly improved by the combination approach. However, regarding power 'globaltest' alone performs best. Consequently, combined testing gains power via the information present in the gene set while it gains 'security' (fewer false positives) by incorporation of miRNA-expression. Often it will even outperform both with regards to APR. This is a recurring result of this simulation which will reappear often. Concerning the more or less acceptable FDR control and the power gain, the 'GlobalAncova' and 'ROAST'-based procedures yielded a very similar result. For the first, however, strange effect occurred regarding the power of the gene set test alone. After rising monotonously with the effect applied, for higher values of δ the power of the gene set test appears to decline again. Under this respect, it is interesting to see that the combination approach does not suffer from this problem. This example for a situation where a gene set test is outperformed
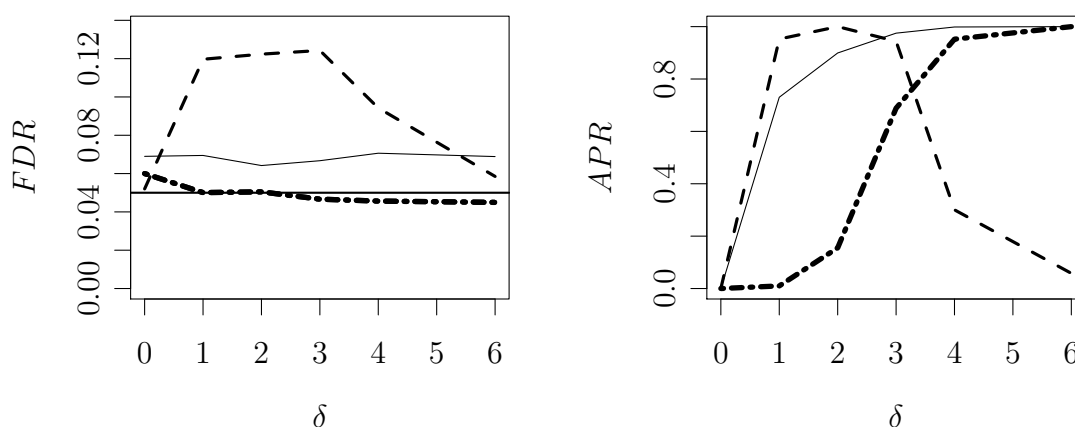

Figure 3: Simulation performance of 'GlobalAncova'. Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

by the combination approach is depicted in Fig. 3. Only 'RepeatedHighDim', as the final self-contained test investigated, also failed to control the FDR level, especially under the null hypothesis. The combined test, yet performing slightly better, mostly ranged at FDR values of around 18 %. As far as the APR is concerned, however, it was quite high for the gene set test.

For the block covariance structure, the global tests kept their FDR level better. The 'globaltest' was even a bit conservative before the combination. 'RepeatedHighDim', however, failed again to control the FDR. As an example regard the 'GlobalAncova' simulation results in Fig. 4. Under this covariance structure again the 'GlobalAncova' on its own had a very low APR. This is the same effect as under autoregressive covariance and is again compensated after the combination. Apart from that, similar results regarding power can be seen as in the case of autoregressive covariance structure.

Finally, the self-contained tests were applied on data with undefined covariance structure. Here the FDR was controlled by the 'globaltest' and 'ROAST' and the combination approaches based on them. For 'GlobalAncova', FDR was not controlled by the test on
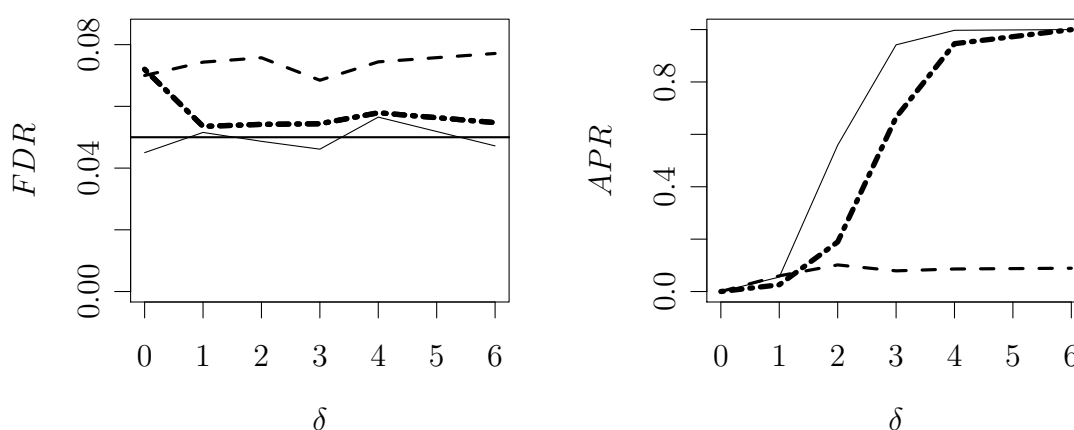

Figure 4: Simulation performance of the 'GlobalAncova'-test for block covariance structure. Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

its own, but by the combined approach. 'RepeatedHighDim' did not control it, neither on its own, nor after the combination with miRNA-wise information. As far as power rates are concerned, they were, as described for other covariance structures, acceptable for all tests but 'GlobalAncova' before the combination.

The FDR was controlled by all competitive gene set tests. Their FDR even tended to be too low so that one can expect them to lose power by that. This was, apart for the Fisher-test-based procedure, lessened by the combination. For an example see the Kolm. Smirnov test in Fig. 5. Regarding the APR, both the gene set tests as well as miRNA-wise testing were outperformed by the combination approach. The only exception is 'Romer', where combined testing yielded approximately the same power as the gene set test. Both had a higher APR than miRNA-wise testing, however.

In the block structure case the competitive tests showed a similar behaviour as for data with autoregressive covariance. The Kolm. Smirnov and the Wilcoxon-test controlled the FDR and had an acceptable power. Again the APR of the Fisher-test was rather low, yet it was improved by the combination. For 'Romer', interestingly, the power was worsened
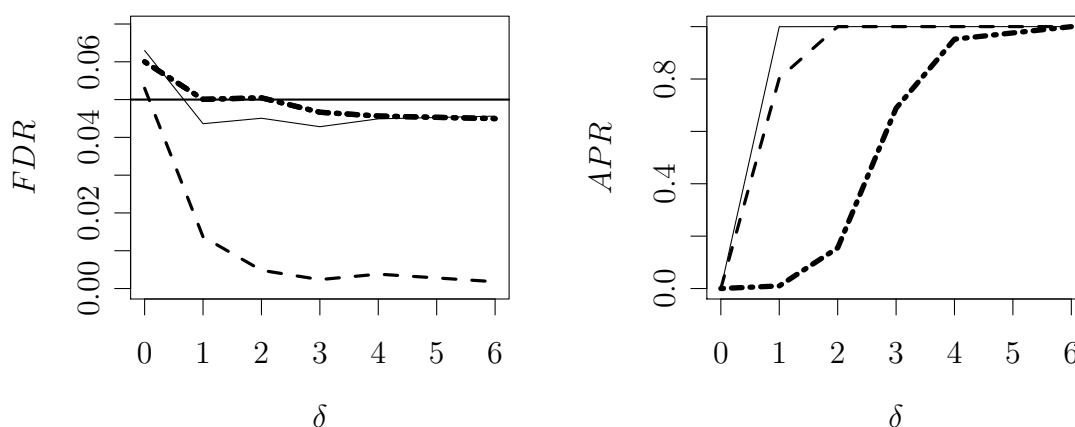

Figure 5: Simulation performance of the Kolm. Smirnov-test. Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

by its application on data with a covariance matrix with block structure (see Fig. 6). Similarly to the case of 'GlobalAncova' described above, this was compensated by the combination approach which again had an acceptable power rate.

When the competitive tests were applied on data with unstructured covariance similar results were yielded as described for the block structure above.

The power rates of the tests applied on autoregressive data are compared in Fig. 7. For other covariance structures see Section 1.3.2 in the Appendix.

All in all, power rates of the different tests on data with autoregressive covariance structure were rather comparable. For the self-contained tests using 'RepeatedHighDim' performed best, 'globaltest' and 'GlobalAncova' did slightly worse and 'Romer' showed a comparatively conservative behaviour. For the competitive tests, Kolm. Smirnov and Fisher had a very high power, while for 'Romer' it was already lower and for the Fisher-test already quite low. Comparing all tests applied, 'KS' and Wilcoxon performed slightly better than 'RepeatedHighDim'. Next came the 'globaltest' and 'GlobalAncova'-based
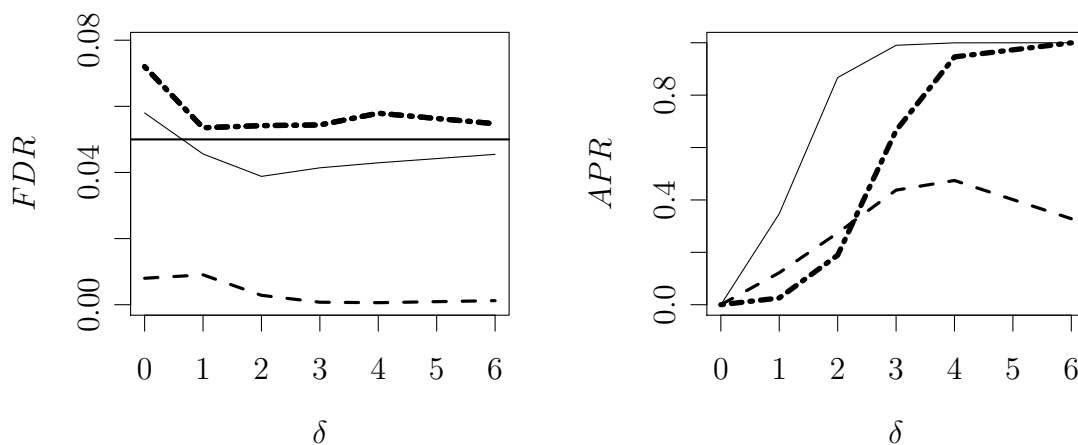

Figure 6: Simulation performance of the 'Romer'-test under block structure covariance of mRNAs and miRNAs. Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

procedures and finally the rotation tests, where 'Romer' was clearly the most conservative one.

For block covariance structure and data with an unstructured covariance matrix most tests performed worse, as is exemplified in Fig. 8. While the APR of the competitive
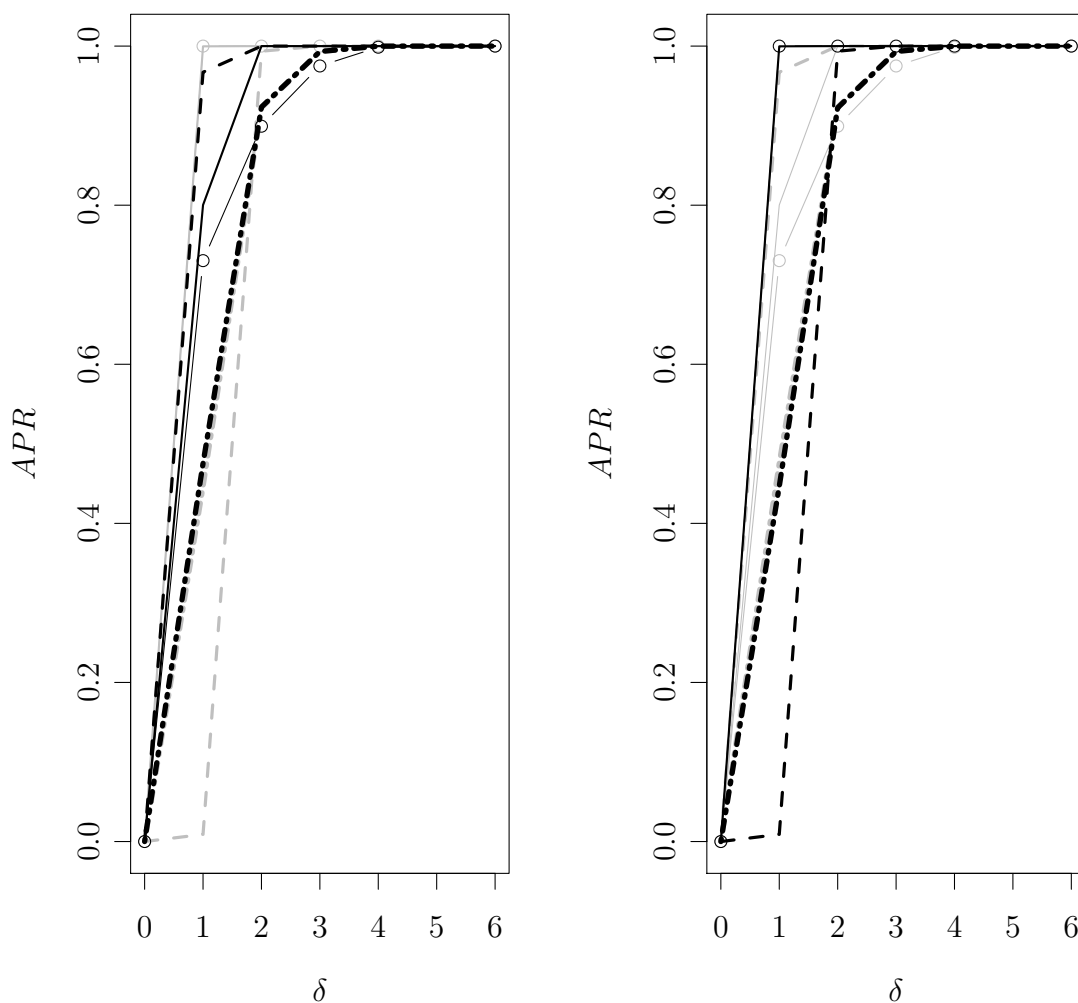

Figure 7: Simulation performance regarding average power rate under autoregressive covariance of mRNAs and miRNAs. On the left self-contained tests are depicted in black, competitive ones in grey, on the right vice versa. All rates depicted after combination. Solid lines 'globaltest'- and Kolm. Smirnov, solid lines with circles 'GlobalAncova'- and Wilcoxon-, dashed lines 'RepeatedHighDim' and Fisher-, dashed and pointed lines 'ROAST'- and 'Romer'-based procedures.

tests remained mostly unharmed, the power of the self-contained tests was significantly lower. The results for block covariance structure and unstructured covariance were mostly identical.
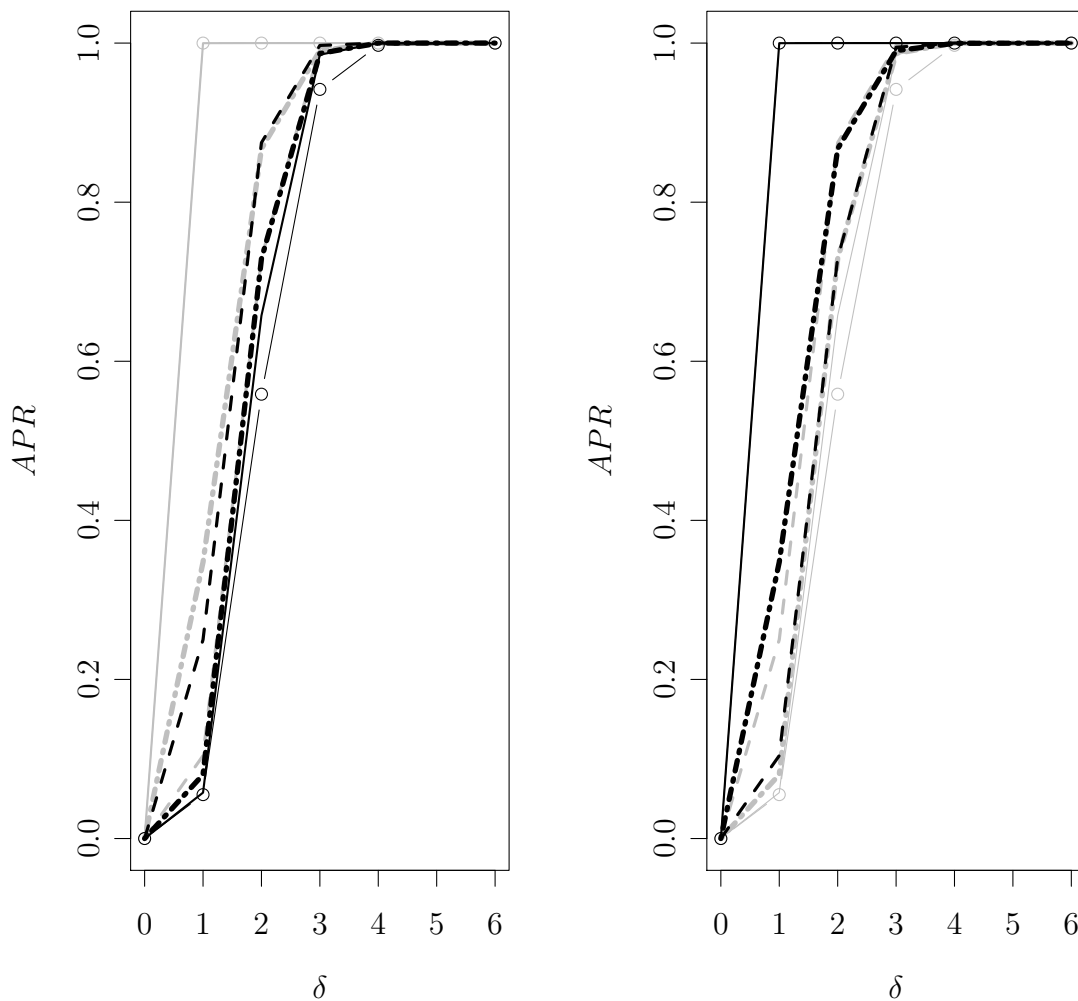
Figure 8: Simulation performance regarding average power rate under block covariance of mRNAs and miRNAs. On the left self-contained tests are depicted in black, competitive ones in grey, on the right vice versa. All rates depicted after combination. Solid lines 'globaltest'- and Kolm. Smirnov, solid lines with circles 'GlobalAncova'- and Wilcoxon-, dashed lines 'RepeatedHighDim' and Fisher-, dashed and pointed lines 'ROAST'- and 'Romer'-based procedures.

3.1.2 Non-overlapping Gene Sets and strong Gene Set Effect

In the second simulation conducted the effect of miRNAs on Gene Sets was stronger. This intended to test which combination tests can withstand extreme differential regulation.

For the global tests under autoregressive covariance, the result here was similar to that in the simulation before. They did not really control the FDR, even though the situation was improved after the combination. Despite the strong effect put on the gene sets, 'GlobalAncova' had practically no power, as can be seen in Fig. 9. Interestingly, this is worse than in the simulation before. For a possible explanation of this behaviour, see Section 4.1, page 52 in the Discussion. Nevertheless, the effect was again compensated by the combination approach. For the other self-contained tests the APR was maximal. After the combination power losses were to be seen for 'ROAST'.

For data with unstructured- or block structured covariance self-contained test performance was similar. Combined testing, however, now enabled the 'globaltest' and 'GlobalAncova' test, after incorporation of miRNA expression, to control the FDR.

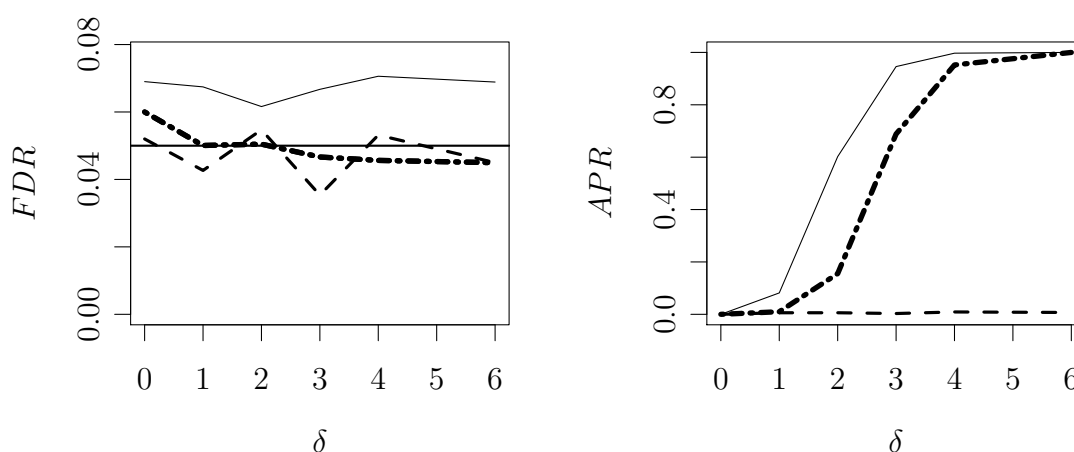**Figure 9: Simulation performance of the 'GlobalAncova'-test under autoregressive structure covariance of mRNAs and miRNAs.** Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

As far as the competitive tests are concerned, they all controlled the FDR under autoregressive covariance. They all were rather conservative. The most extreme example is the Fisher Test depicted in Fig. 10. After the combination the behaviour was less conservative. For the enrichment tests, however, this does not lead to an increased APR. Indeed, this was impossible as it jumped to 1 immediately once $\delta > 0$. On the contrary, for 'Romer' the combination approach had a higher FDR (which was still under control, however) yet also a lower power than the gene set test on its own.

As far as other covariance structures are concerned, the enrichment tests' performance very much resembled the autoregressive case. As far as 'Romer' is concerned, however, a power loss was found for larger values of δ which was compensated by the combination.

When comparing the power rates of the combination tests, most achieve the maximal power, i. e. have an APR of 0 under the complete null, and and APR of 1 otherwise. As can be seen in Fig. 11, only the rotation tests and 'GlobalAncova' do not reach such a high power rate.
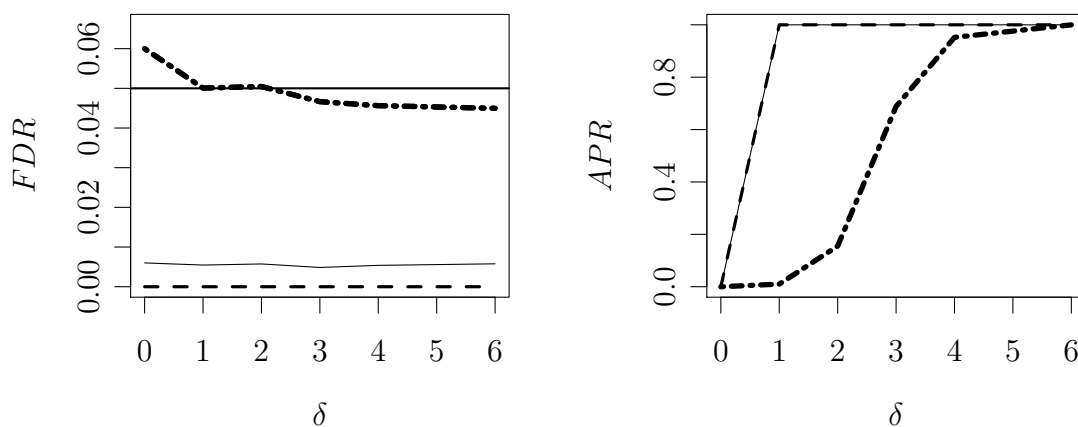
Figure 10: Simulation performance of the Fisher-test under autoregressive structure covariance of mRNAs and miRNAs. Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

For the other two covariance structures similar results were found, with the exception of 'Romer' which, as mentioned above, lost much power on its own. This, however, was compensated in the combined testing (see Fig. 12). Nevertheless, it dragged the combination approach as well, especially for larger values of δ, as can be seen in Fig. 12.
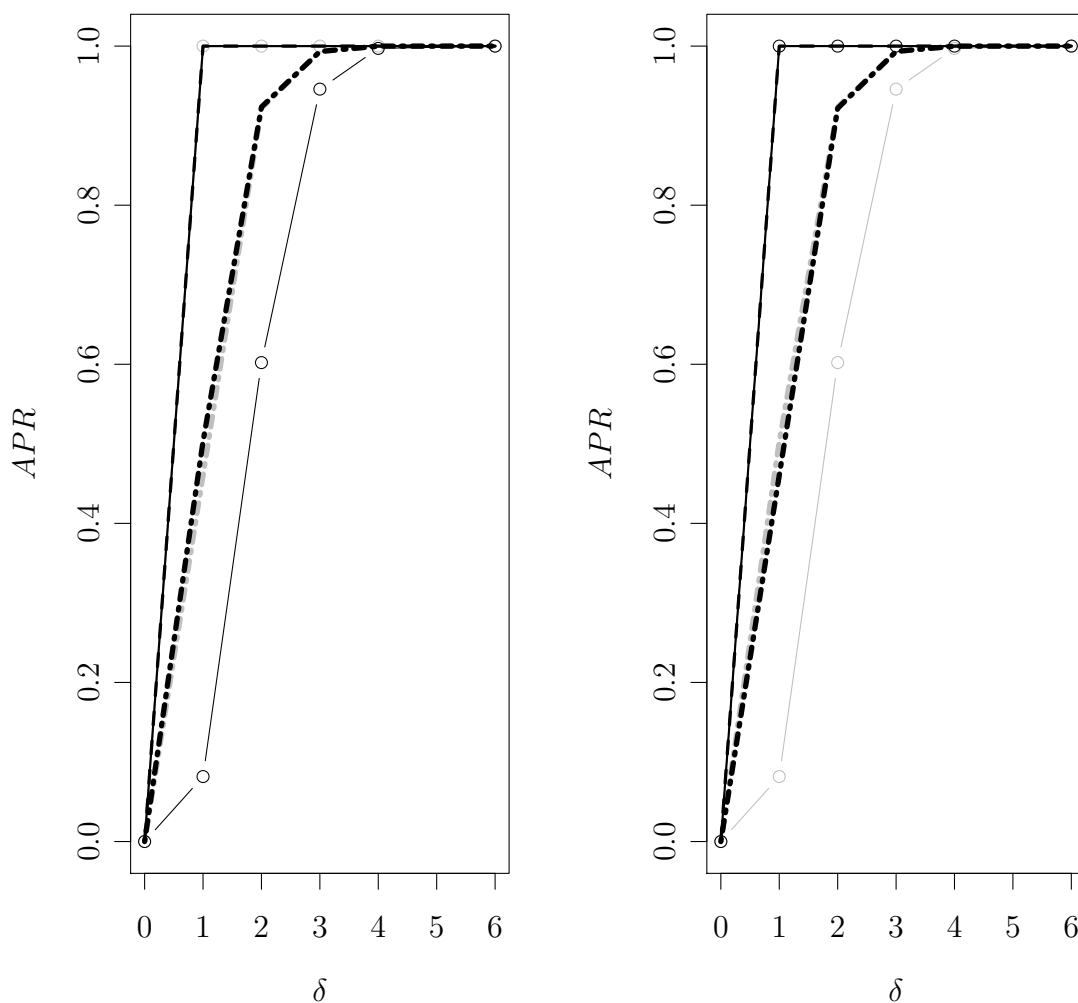
Figure 11: Simulation performance regarding average power rate under autoregressive structure covariance of mRNAs and miRNAs. On the left self-contained tests are depicted in black, competitive ones in grey, on the right vice versa. All rates depicted after combination. Solid lines 'globaltest'- and Kolm. Smirnov, solid lines with circles 'GlobalAncova'- and Wilcoxon-, dashed lines 'RepeatedHighDim' and Fisher-, dashed and pointed lines 'ROAST'- and 'Romer'-based procedures.

See Section 4.1, page 52 for possible explanations of this behaviour of the gene set test.

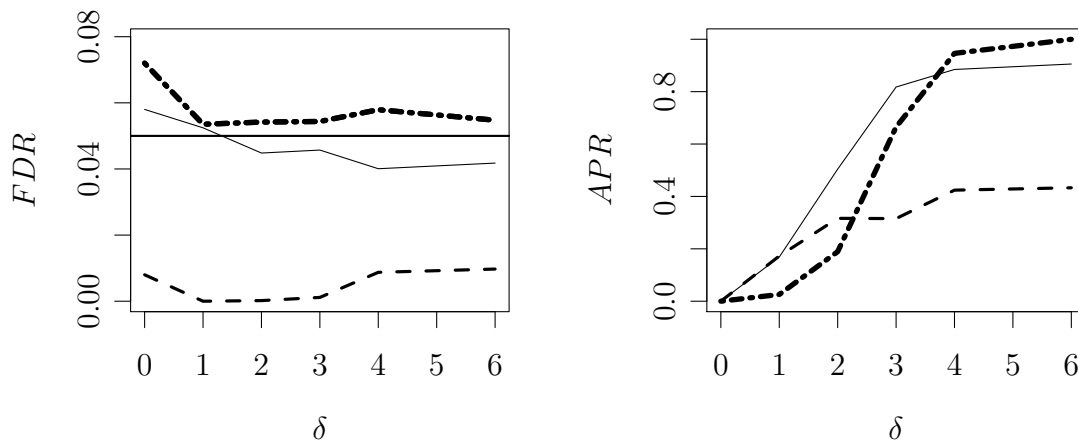**Figure 12: Simulation performance of the 'Romer'-test under autoregressive structure covariance of mRNAs and miRNAs.** Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

3.1.3 Overlapping Gene Sets and Weak Gene Set Effect

Here, the more realistic assumptions of randomly overlapping gene sets was introduced. The expectation and variance of gene set size and overlap were chosen in such a way as to reflect empirical data taken from the 'microCosm' database. There, around 700 miRNAs were present with overlaps of about 10 % between two given miRNAs. The simulation parameters resembled these parameters on a smaller scale (see Section 2.6).

Under the strong dependence introduced between the gene set statistics by the overlaps the already liberal global tests did not nearly control the FDR, especially when the effect $\delta > 0$. The combination approach could not lessen this problem (see Fig. 13 for an example). The same is true for 'ROAST' which displays an FDR proportional to the effect δ applied. The FDR control of 'ROAST', yet still way too liberal, might be considered best of the self-contained tests applied.

A similar result was obtained for data with block covariance structure. Surprisingly, however, the FDR of 'ROAST' is controlled by the combination approach, the power of

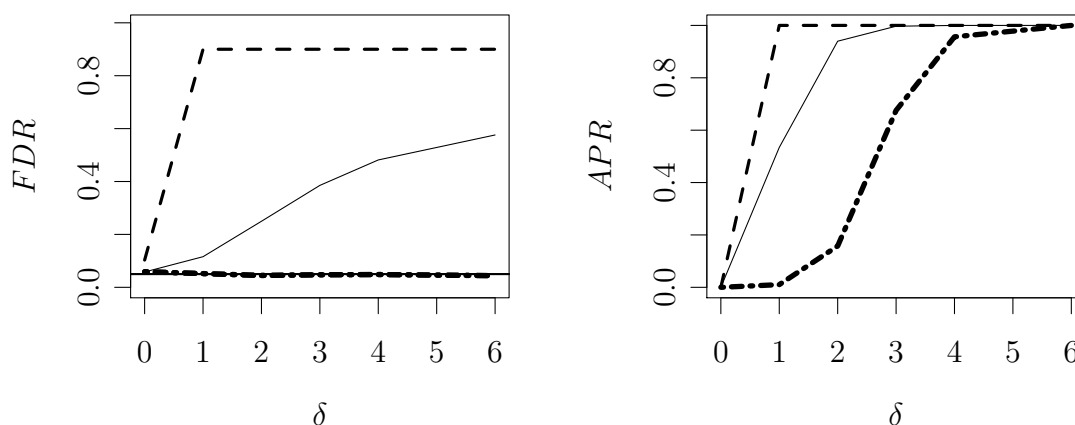**Figure 13: Simulation performance of the 'ROAST'-test under block structure covariance of mRNAs and miRNAs.** Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

which still ranges close to the one of the gene set test alone. After application on data with unstructured covariance matrix the self-contained tests performed comparably to the case with block covariance.

The competitive tests, on the contrary, were, as in the simulations before, able to control the FDR. All but the Wilcoxon test even had an a little too low FDR, while all of them had a rather large power. The combination tests based on the Kolm. Smirnov and the Wilcoxon test had the same power as their respective gene set tests. It was 0 for the null and 1 for the alternative. The remaining two tests also had high APRs, which, however, were a little lower than those of the corresponding gene set tests.

As far as the power rate is concerned, it was extremely high for the self-contained tests apart from 'ROAST' reaching the maximum value immediately under the alternative. For 'ROAST' the power curve was a little lower, yet it was still absolutely acceptable.

Also when the synthetic data followed a normal distribution with a covariance matrix of block structure, the competitive tests were able to control the false discovery rate. The
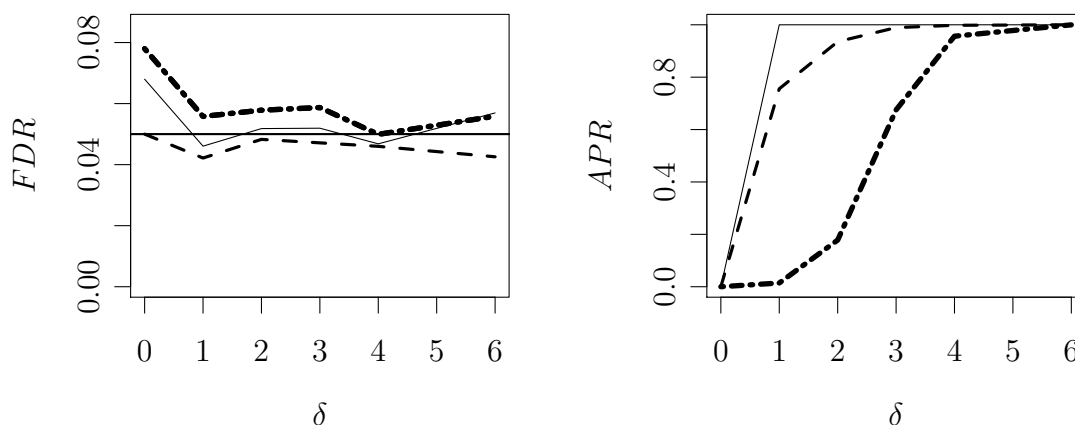
Figure 14: Simulation performance of the Wilcoxon-test under unstructured covariance of mRNAs and miRNAs. Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

power rate soon reached its maximal value for the gene set tests alone and with a little higher values of δ the power rate of the combined tests did the same. The only exception was once more the 'Romer' test which had a very low APR before the combination which was shifted to quite acceptable values afterwards. Similar results were obtained with
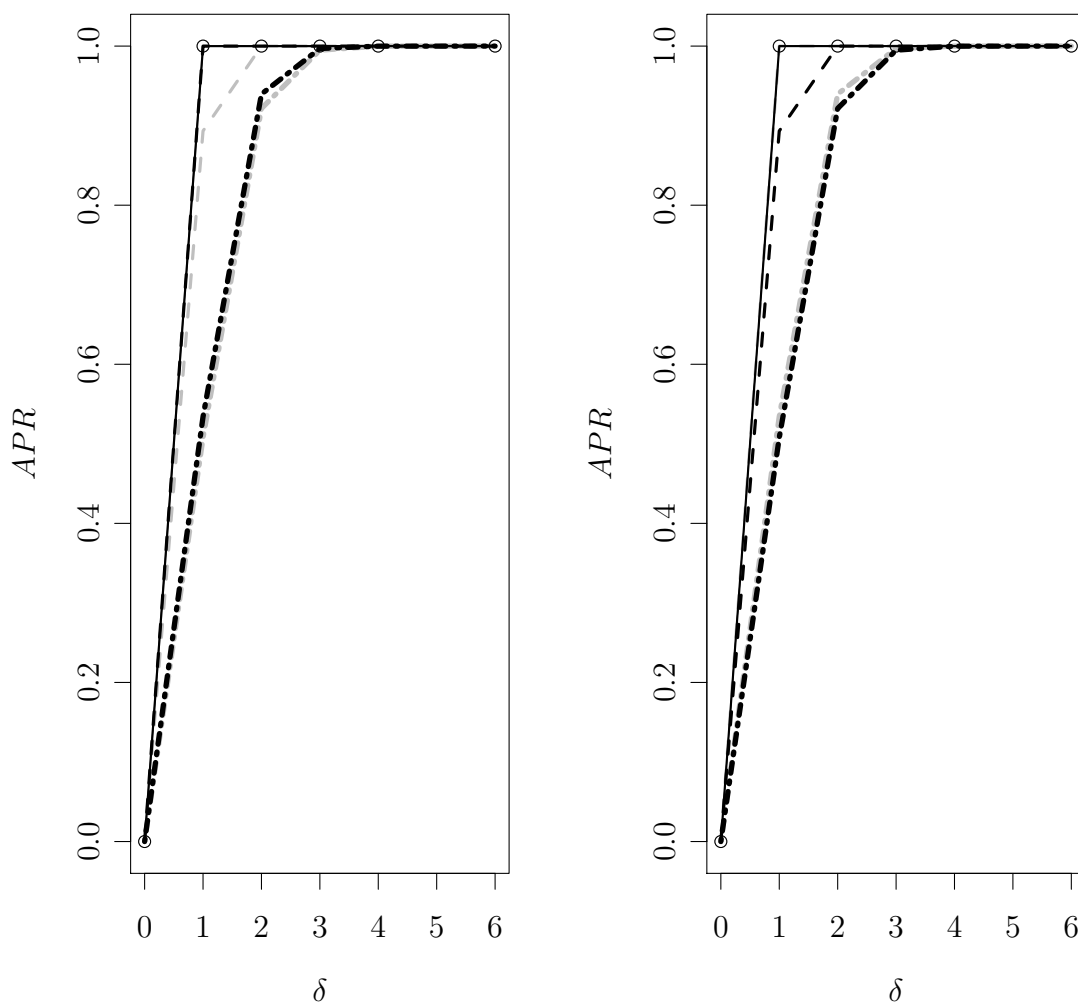
Figure 15: Simulation performance regarding average power rate under autoregressive covariance of mRNAs and miRNAs. On the left self-contained tests are depicted in black, competitive ones in grey, on the right vice versa. All rates depicted after combination. Solid lines 'globaltest'- and Kolm. Smirnov, solid lines with circles 'GlobalAncova'- and Wilcoxon-, dashed lines 'RepeatedHighDim' and Fisher-, dashed and pointed lines 'ROAST'- and 'Romer'-based procedures.

data of unstructured covariance. As an example, regard the FDR and APR for the Wilcoxon-based test on data of unstructured covariance in Fig. 14.

The performance regarding power of the different combined tests is depicted in Fig. 15. The APR of most of the global tests reaches the maximum value of 1 once $\delta > 0$. The
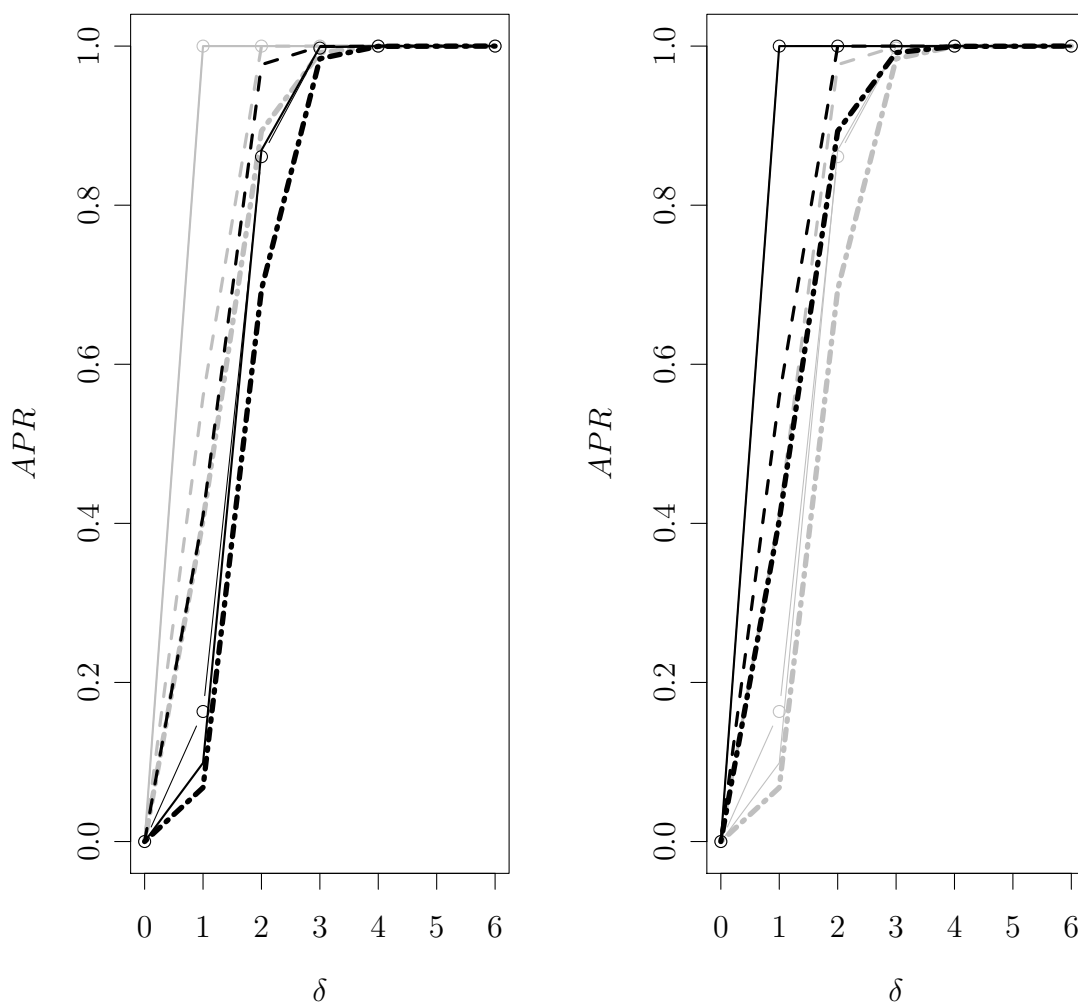
Figure 16: Simulation performance regarding average power rate under block structure covariance of mRNAs and miRNAs. On the left self-contained tests are depicted in black, competitive ones in grey, on the right vice versa. All rates depicted after combination. Solid lines 'globaltest'- and Kolm. Smirnov, solid lines with circles 'GlobalAncova'- and Wilcoxon-, dashed lines 'RepeatedHighDim' and Fisher-, dashed and pointed lines 'ROAST'- and 'Romer'-based procedures.

same was observed for the non-rotating enrichment tests. Only the Fisher test-based procedure has a slightly lower APR for $\delta = 1$. The power rate of 'ROAST' and 'Romer', however, was acceptable yet considerably smaller than for the other tests. Only for $\delta \geq 3$ it reached the maximum value. This is approximately the same power curve as it was observed in the first simulation with independent gene sets and a weak gene set effect (see Section 3.1.1).

The tests led to quite different results on the other two covariance structures than described above. As an example regard the power curves for the block covariance structure in Fig. 16. The plot for unstructured data is given in Section 4.3 of the Appendix.

The power rates of all self-contained tests declined, most drastically for 'globaltest' and 'GlobalAncova' which were overtaken by 'RepeatedHighDim'. 'ROAST' remained to be the worst test concerning power. For the competitive tests the effect was less dramatic. The Kolm. Smirnov, Wilcoxon and 'Romer'-based procedures kept their high power, while for the Fisher Test-based one it decreased. Nevertheless, basing the combination approach on 'Romer' led to the worst result in competitive testing as far as the power rate is concerned. Apart from 'Romer' the competitive tests had a higher power than all the self-contained ones applied. For data with unstructured covariance a very similar result was gained.

3.1.4 Overlapping Gene Sets and Strong Gene Set Effect

The intention of this simulation was to test the procedures' FDR control under very harsh conditions. Both challenges from the two simulation before were combined here: a very strong gene set effect and gene set overlaps.

As can be expected, FDR control for the self-contained tests worsened here as compared to the simulations before. Therefore, the results of these tests shall not be detailed further. For the plots of their performance see Section 4.3 in the Appendix.

Emphasis, on the other hand, shall be put on the competitive tests. They performed quite well throughout the simulations before and the interest is to see whether this was still the case here.

For data with autoregressive covariance structure, the FDR was maintained by all four combination approaches that were based on a competitive gene set test. This was also the case for the Kolm. Smirnov, Wilcoxon and Fisher gene set test. 'Romer' did not maintain the 5 % level, as can be seen in Fig. 17. The APR was maximal for the three
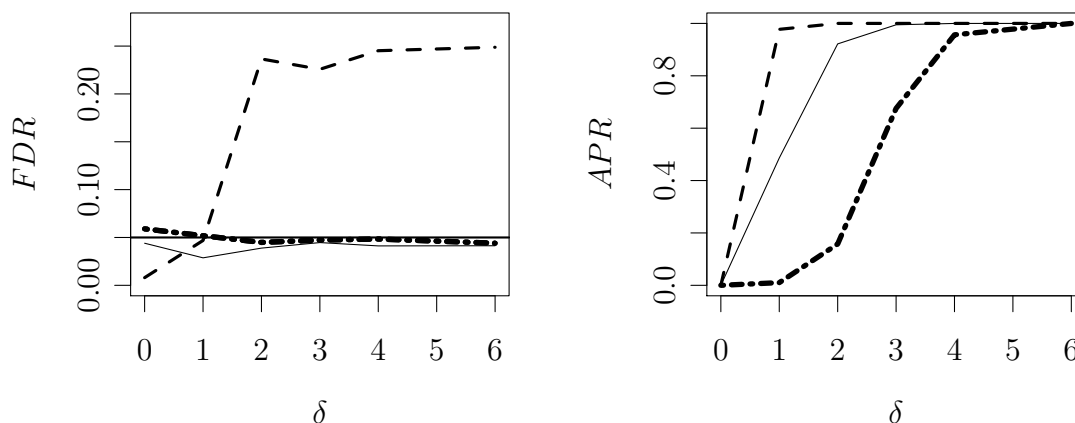
Figure 17: Simulation performance of the 'Romer'-based test under autoregressive covariance of mRNAs and miRNAs. Against the effect size δ FDR on the left and APR on the right are plotted. The 5 % FDR threshold is indicated by a horizontal line. Dashed and pointed line for miRNA-wise testing, dashed line for Gene Set Test and solid line for combination.

aforementioned tests, while for 'Romer' it was worse after the combination, rendering it a rather conservative test under autoregressive covariance.

When applied on expression data of block structured covariance, a similar result appeared. 'Romer', however, had a much worse power, also dragging the APR of the com-
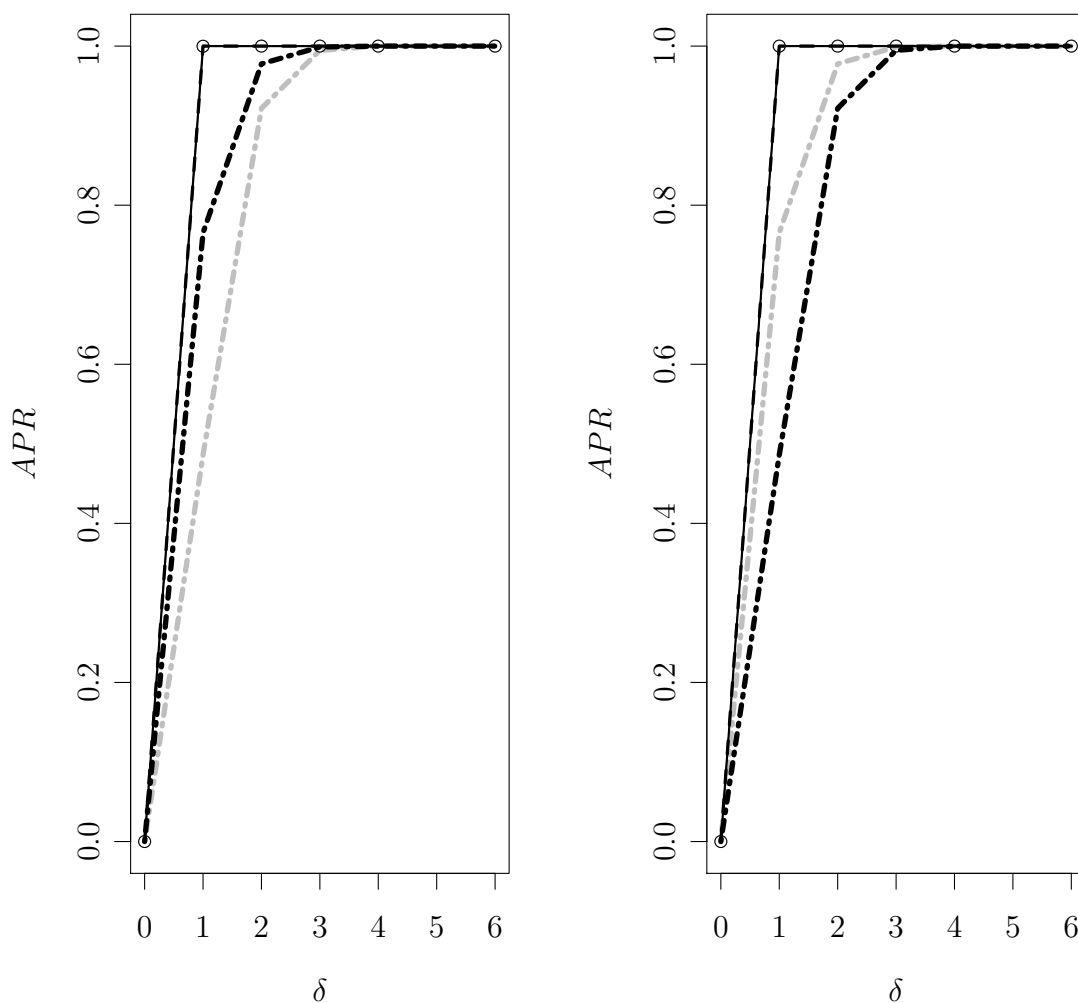
Figure 18: Simulation performance regarding average power rate under autoregressive structure covariance of mRNAs and miRNAs. On the left self-contained tests are depicted in black, competitive ones in grey, on the right vice versa. All rates depicted after combination. Solid lines 'globaltest'- and Kolm. Smirnov, solid lines with circles 'GlobalAncova'- and Wilcoxon-, dashed lines 'RepeatedHighDim' and Fisher-, dashed and pointed lines 'ROAST'- and 'Romer'-based procedures.

bination approach a little. Nevertheless, combining miRNA- and mRNA-data allowed to compensate for this. For data having an unstructured covariance matrix practically the same result was found.

A comparison of the average power rates of all procedures applied under autoregressive covariance structure is given in Fig. 18. It was merely the same for other covariance structures. All tests had the maximally achievable APR apart from the rotation tests. These, however, also had considerably high power rates.

3.1.5 Summary

To help the reader to cope with the many simulation results acquired, this chapter intends to briefly sum up the findings of the simulations elaborated above. An interpretation and discussion will be given in Section 4.1.

First, it was shown that FDR control depends on the gene set test, as the linear models of Smyth (2004) always do so. It was not guaranteed when self-contained tests were used. As can be seen in Tab. 2, they exceeded the level sought to be controlled quite often. The combination approach managed to alleviate this behaviour but could not remedy it entirely. This is especially evident in the case that gene sets were modelled to be overlapping and/or strong gene set effects were applied. Indeed, once gene sets overlapped only the combination approaches based on the 'ROAST' rotation test were able to maintain the FDR level, and this just so long as the gene set effect was weak.

Second, competitive gene set tests controlled the FDR under all circumstances. Many competitive tests were rather conservative with unnecessarily few type I errors. The only exception to that is 'Romer', which under overlapping gene sets and a strong gene set effect exceeded the 5% FDR threshold. The combination approaches, however, alleviated both problems: They always were in control if the FDR. Additionally, the combination also lessened the conservativeness of many tests.

Next, it was of interest to see how the combination influenced the APR of the tests. Under all simulations performed the power of miRNA-wise testing was outperformed by the combination approach. In many situations this was also true for the corresponding

gene set test on its own, leading to the combined testing approach to outperform both tests. Under non-overlapping gene sets combination testing almost always had the highest APR. Under a strong gene set effect, the rotation tests gained in power and overtook the combination approach.

As far as the simulations with overlapping gene sets are concerned, for the weak effect the combination approach outperformed almost always the tests it is based upon for the competitive tests. For the self-contained tests this was only the case for RepeatedHighDim and 'ROAST'. Finally, under a strong effect again only the rotation tests outperformed their combined counterparts.

Table 2: *FDR control in Simulations.* Simulations had non-overlapping or overlapping Gene Sets (GS) and strong or weak gene set effects. Results are grouped according to the gene set test applied (self-contained above, competitive below) and the data's covariance structure. *FDR* control was conservative (c), slightly conservative (s. c), maintained (m), slightly exceeded (**s. e**) or exceeded (**e**).

GS Test applied	Cov. Structure	Gene Set Test alone				Combined Testing			
		Non-o. GS		Overl. GS		Non-o. GS		Overl. GS	
		Weak	**Strong**	**Weak**	**Strong**	**Weak**	**Strong**	**Weak**	**Strong**
globaltest	Autoregressive	e	e	e	e	e	e	e	e
	Block	m	m	e	e	m	m	e	e
	Unstruc.	m	c	e	e	m	c	e	e
GlobalAncova	Autoregressive	e	m	e	e	s. e	m	e	e
	Block	e	s. e	e	e	m	m	e	e
	Unstruc.	e	e	e	e	m	m	e	e
RepeatedHighDim	Autoregressive	e	e	e	e	e	e	e	e
	Block	e	e	e	e	e	e	e	e
	Unstruc.	e	e	e	e	e	e	e	e
ROAST	Autoregressive	s. e	s. e	e	e	m	m	e	e
	Block	c	s. c	e	e	m	m	m	e
	Unstruc.	c	c	e	e	m	m	m	e
Kolm. Smirnov	Autoregressive	c	c	c	c	m	m	m	m
	Block	c	c	s. c	c	m	m	m	m
	Unstruc.	c	c	s. c	c	m	m	m	m
Wilcoxon	Autoregressive	c	c	m	m	c	s. c	m	m
	Block	c	c	m	m	m	m	m	m
	Unstruc.	c	c	m	m	m	s. c	m	m
Fisher	Autoregressive	c	c	c	c	c	c	c	c
	Block	c	c	c	c	c	c	c	c
	Unstruc.	c	c	c	c	c	c	c	c
Romer	Autoregressive	c	c	c	e	s. c	m	s. c	m
	Block	c	c	c	e	m	m	m	m
	Unstruc.	c	c	c	e	m	m	m	m

Table 3: *APR* **in Simulations.** Simulations had non-overlapping or overlapping Gene Sets (GS) and strong or weak gene set effects. Results are grouped according to the gene set test applied (self-contained above, competitive below) and the data's covariance structure. Powers are compared of miRNA-wise (m), gene set (s) and combined ($\mathbf{C}$) testing. $<, \leq, \sim$ stand for lower, lower to equal and comparable power, respectively.

GS Test in comb. Testing	Cov. Struc.	Non-o. GS Weak	Non-o. GS Strong	Overl. GS Weak	Overl. GS Strong
globaltest	Autoregr.	m < s $\leq$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**
	Block	m < s < **C**	m < s $\sim$ **C**	m < **C** $\sim$ s	m < s $\sim$ **C**
	Unstruc.	m < s < **C**	m < s $\sim$ **C**	m < **C** $\sim$ s	m < s $\sim$ **C**
GlobalAncova	Autoregr.	s < m < **C**	s < m < **C**	m < s $\sim$ **C**	m < s $\sim$ **C**
	Block	s < m < **C**	s < m < **C**	m < **C** $\sim$ s	m < s $\sim$ **C**
	Unstruc.	s < m < **C**	s < m < **C**	m < **C** < s	m < s $\sim$ **C**
RepHighDim	Autoregr.	m < s $\leq$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**
	Block	m < s $\leq$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**
	Unstruc.	m < s $\leq$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**
ROAST	Autoregr.	m < **C** < s	m < **C** < s	m < **C** < s	m < **C** < s
	Block	m $\leq$ s < **C**	m < **C** < s	m < s $\sim$ **C**	m < **C** < s
	Unstruc.	m $\leq$ s < **C**	m < **C** < s	m < s $\sim$ **C**	m < **C** < s
Kolm. Smirnov	Autoregr.	m < s < **C**	m < s $\sim$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**
	Block	m < s < **C**	m < s $\sim$ **C**	m < s < **C**	m < s $\sim$ **C**
	Unstruc.	m < s < **C**	m < s $\sim$ **C**	m < s < **C**	m < s $\sim$ **C**
Wilcoxon	Autoregr.	m < s < **C**	m < s $\sim$ **C**	m < s $\sim$ **C**	m < s $\sim$ **C**
	Block	m < s < **C**	m < s $\sim$ **C**	m < s < **C**	m < s $\sim$ **C**
	Unstruc.	m < s < **C**	m < s $\sim$ **C**	m < s < **C**	m < s $\sim$ **C**
Fisher	Autoregr.	m < s $\sim$ **C**	m < s $\sim$ **C**	m < s < **C**	m < s $\sim$ **C**
	Block	m < s $\leq$ **C**	m < s $\sim$ **C**	m < s < **C**	m < s $\sim$ **C**
	Unstruc.	m < s $\leq$ **C**	m < s $\sim$ **C**	m < s < **C**	m < s $\sim$ **C**
Romer	Autoregr.	m < s $\leq$ **C**	m < **C** < s	m < **C** < s	m < **C** < s
	Block	s < m < **C**	s < m $\leq$ **C**	s < m < **C**	s $\leq$ m < **C**
	Unstruc.	s < m < **C**	s < m $\leq$ **C**	s < m < **C**	s $\leq$ m < **C**

3.2 Rat Neurogenesis Data

The analysis of the data of Nielsen *et al.* (2009) will be described in the following. Focus will be put on the comparison of the authors' original analysis and the application of the combination approaches described above. This chapter will therefore start with the miRNAs that in the original publication by Nielsen *et al.* (2009) had been found to be significantly differentially regulated with a fold change larger than two. These include 31 miRNAs, 20 of which were down-regulated in E13. All of the up-regulated ones had

been validated in RT-PCR by Nielsen *et al.* (2009). In a final chapter the miRNAs not mentioned in the original publication but detected by some or all combination approaches will be discussed.

3.2.1   Originally detected miRNAs

If the 31 miRNAs are subdivided into groups according to their results in the original and in the re-analysis, one arrives at Tab. 4. 22 miRNAs were detected by all of the combination tests (uppermost group in Tab. 4). The first six had already been described by Nielsen *et al.* (2009) as being up-regulated with inversely correlated target genes. Note that their definition of inverse correlation was that the subset of target genes regulated in the inverse direction of the miRNA was significant according to their two-sided Fisher test. Five miRNAs that were detected by all combination tests applied were down-regulated and had been mentioned as having inversely regulated gene sets. Apart from that, six other down-regulated miRNAs had not had any of their two subsets called significant, but for them the *p*-value for the up-regulated subset was lower than that for the down-regulated one. Of the remaining five miRNAs, two were down-regulated with a stronger indication of target set down-regulation but without significance in the target subsets, one was down-regulated but had both target subsets significant with the lower *p*-value in the up-regulated one, and two were up-regulated without significant gene sets but stronger significance in the up-regulated subsets.

Furthermore, there were 5 miRNAs which were detected by all procedures but those based on rotation tests. Out of these, miR-218, which showed an increase in its fold change, had already been described by the authors as having a differential up-regulated subset of target genes. The down-regulated one, on the other hand, was not significant. For miR-7, which was also among these five miRNAs, Nielsen *et al.* (2009) had found neither subset of the target genes to be differential. miR-199a was down-regulated and had a differential up-regulated target set. However, as the authors mention this miRNA had far fewer up-regulated targets than expected and was therefore already by them not counted as having an inversely correlated target set. Another down-regulated miRNA

not detected by the rotation tests, miR-214, had no differentially expressed target gene subsets. It was not detected even though its up-regulated targets were more significant than its down-regulated ones. Finally, miR-124a falls into this group. It is an example of the rotation tests' conservative behaviour: even though it was differentially up-regulated and had been shown to have an inversely correlated gene set, it was not detected by them.

Three more miRNAs that had been found by Nielsen *et al.* (2009) were detected by all test but one of the rotation tests. One of them, the down-regulated miR-126, was not significant in 'ROAST'-based testing. Its gene set was not significant in either direction, but its down-regulated subset had been assigned a lower p-value by Nielsen *et al.* (2009). miR-210 and miR-19a were not detected by the 'romer'-based procedure. miR-210 was another down-regulated miRNA whose target set was not significant in either direction, but with a lower p-value for the up-regulated subset. miR-19a was down-regulated as well. However, in the original publication it had been found to have a gene set just reaching the threshold of being inversely regulated. miR-290, finally, was a down-regulated miRNA whose down-regulated subset was significant in Nielsen *et al.* (2009).

3.2.2   Additionally detected miRNAs

Further miRNAs were detected by all or some of the combined testing procedures which had not been described in Nielsen *et al.* (2009). For their complete list see Section 4.3 in the Appendix. Here, just their numbers shall be listed, which already allows an estimation of the liberality of the tests applied.

There were eighteen further miRNAs that were significant in miRNA-wise testing as well as in all combined tests applied. Apart from these, further miRNAs had strong enough effects in their target sets so that they were detected even though miRNA-wise testing did not call their expression significantly differential.

The 'Romer'-based procedure, as the most conservative one, detected three further miRNAs. Next came the test based on the second rotation test, 'ROAST'. It already detected 25 further miRNAs (additionally to the eighteen described above).

The tests based on the enrichment tests, the Kolmogorov Smirnov, Wilcoxon and

Table 4: Comparison of Results from Nielsen *et al.* (2009) and re-Analysis.
For each of the miRNAs identified by Nielsen *et al.* (2009), the regulation in E13 (Reg. In E13), the originally published result for the up- and down-regulated subsets of the gene sets (Original Gene Set) and the results of the re-analysis based on Global Tests (GTs), enrichment tests (ETs) and rotation tests (RTs) are given. A test was result was called significant (**sign.**) with a $q < 5\%$, else it was not significant (n. s.). miRNAs are grouped according to result in re-analysis.

miRNA	Reg. In E13	Orig. GS Res. Up	Down	Re-Analysis GTs	ETs	RTs
miR-99a/b	up	n. s.	sig.	**sig.**	**sig.**	**sig.**
miR-9	up	n. s.	sig.	**sig.**	**sig.**	**sig.**
miR-100	up	n. s.	sig.	**sig.**	**sig.**	**sig.**
miR-181b/c	up	n. s.	sig.	**sig.**	**sig.**	**sig.**
miR-125a/b	up	n. s., up > down		**sig.**	**sig.**	**sig.**
miR-222	down	sig.	n. s.	**sig.**	**sig.**	**sig.**
miR-291-3p	down	sig.	n. s.	**sig.**	**sig.**	**sig.**
miR-92	down	sig.	n. s.	**sig.**	**sig.**	**sig.**
miR-145	down	sig.	n. s.	**sig.**	**sig.**	**sig.**
miR-183	down	sig.	n. s.	**sig.**	**sig.**	**sig.**
miR-363-3p	down	sig., up > down		**sig.**	**sig.**	**sig.**
miR-143	down	n. s., up > down		**sig.**	**sig.**	**sig.**
miR-200b/c	down	n. s., up > down		**sig.**	**sig.**	**sig.**
miR-20b	down	n. s., up > down		**sig.**	**sig.**	**sig.**
miR-219	down	n. s., up > down		**sig.**	**sig.**	**sig.**
miR-18	down	n. s., up >> down		**sig.**	**sig.**	**sig.**
miR-205	down	n. s., up < down		**sig.**	**sig.**	**sig.**
miR-292-3p	down	n. s., up < down		**sig.**	**sig.**	**sig.**
miR-218	up	sig.	n. s.	**sig.**	**sig.**	n. s.
miR-7	up	n. s.	n. s.	**sig.**	**sig.**	n. s.
miR-124a	slightly up	n. s.	sig.	**sig.**	**sig.**	n. s.
miR-214	down	n. s., up > down		**sig.**	**sig.**	n. s.
miR-199a	down	few target mRNAs		**sig.**	**sig.**	n. s.
miR-19a	down	slightly sig.	n. s.	**sig.**	sig.	**sig.** (ROAST)) n. s. (Romer)
miR-210	down	n. s.	n. s.	**sig.**	sig.	**sig.** (ROAST)) n. s. (Romer)
miR-126	down	n. s., up < down		**sig.**	sig.	**sig.** (Romer) n. s. (ROAST)
miR-290	down	n. s.	sig.	**sig.**	**sig.** (KS, F) n. s. (W)	n. s.

Fisher Test, further detected 36 miRNAs, 45 miRNAs and 76 miRNAs, respectively.

Finally, procedures based on the global tests 'globaltest', 'GlobalAncova' and 'RepeatedHighDim' assigned significant q-values to nearly all of the miRNAs. They detected all but 31 miRNAs, 34 miRNAs or 35 miRNAs, respectively.

4 Discussion

4.1 Improvements over miRNA-wise and Gene Set Testing in Simulations

The combination approach decreased the error rate of many gene set tests in the simulations. On the one hand, this is just to be expected considering that more information is provided to the combination test than to gene set testing alone. On the other, it is probably also caused by the well defined character of miRNA-wise testing. Testing genes side by side has been done for many years now and the empirical Bayes method of Smyth (2004) has been successfully applied to many microarray data sets already. The multivariate normally distributed data it was applied on in the simulations can be efficiently analysed by the moderated t-statistics. It is therefore not surprising that it led to less false positives than testing gene sets. The latter have to cope with many challenges that do not exist in miRNA-wise testing: correlation of genes, which cannot always be compensated by subsequent multiple testing correction, for example. Gene set overlaps, as has been shown in the simulations, is another major issue. Finally, incomplete target annotation or incorrect target prediction comes into play.

Combining miRNA and gene set information did not only result in better FDR control, but often also in higher *averagepowerrates*. Again, this can be explained with the fact that for this test the most information is present. This led to a good detection rate also when small effects were present. However, the relatively strong effect (even when it was designated as 'weak' to distinguish it from the even stronger effect) put on gene sets needs to be taken into consideration. The miRNA effect on its targets was more or less equal to the effect of the miRNA itself. Naturally, in reality one might not expect every predicted target to be truly influenced in its expression by a miRNA. On the other hand, as miRNAs have catalytic activity, the subset of true targets can be expected to show much stronger differential expression. This may especially influence the result for the rotation tests, as they construct mean statistics for every gene set which probably detect small effects in the entire gene set better than strong ones in selected genes. Consequently, further

simulations might be worthwhile here to investigate their behaviour in such cases.

Another finding was that competitive gene set tests generally performed better than self-contained ones as they were able to control the FDR in almost all situations. Naturally, as more conservative tests they can cope with too strong gene set effects which may be introduced by gene set overlaps or inter-gene correlation better. Additionally, as they compare gene sets with their complement they also 'punish' sets of less significant genes, leading to a list of p values that might sometimes have a bimodal character. This effect, however, is rather small for real microarray data with its large amount of genes and comparably small gene sets.

Note, however, that apart from the rotation tests competitive gene set testing as performed here ignores gene-wise correlation structures. This assumption was not necessarily met in the simulations. That the competitive tests still controlled FDR and had a large power might therefore change with stronger covariance between a larger number of genes than been modelled here. That this might be the case is also indicated by the dependence of the test result on the correlation structure applied. Since allocation of genes to gene sets was resampled in every simulation repetition, the question whether the true cause for the influence the covariance structure had on the results were the different numbers of genes that had a covariance other than zero remains. Again, this leaves room for further simulation studies. They then might also include a logical feature of gene sets that was omitted here, namely that indeed the genes inside a miRNA gene set show a certain grade of correlation.

Interestingly, 'GlobalAncova' lost power in the non-overlapping simulations with larger effects δ or even had none for all δ. For the simulations with overlapping gene sets, were gene sets varied around 120 genes in size, the effect had not been observed. This effect most probably was caused by the fact that the gene sets there contained just 50 mRNAs and therefore were rather small. For 'romer', however, a similar result was obtained for simulations depending on the type of the correlation matrix. It was seen in both the context of non- and of overlapping gene sets. Probably this was the price one has to pay for a test that considers gene-wise correlations, namely that it can get quite conservative

under certain circumstances. Fortunately, on both cases the combination approach had an acceptable power, most often still outperforming miRNA-wise testing.

4.2 Application on real Data

The advantage of a simulation is that one knows which miRNAs really had an effect, so that one can exactly calculate FDR and APR. However, for the reasons mentioned above and because a simulation can never fully model reality, it is important to test procedures on real data sets. This was done here with the data of Nielsen *et al.* (2009). Even though their results have often been taken as a reference when describing the re-analysis of their data one must not forget that they cannot be treated as a gold standard. Indeed, their way of detecting significant gene sets is intuitive but may be flawed as it performs two tests for each miRNA. Nevertheless, most of their results were confirmed in the reanalysis. Besides, they have validated their up-regulated miRNAs.

Quite a lot of miRNAs which one would wish to find were indeed detected. This can be especially said for a subset of miRNAs that had been described and validated by the authors before. It is striking, for example, that most of the miRNAs whose differential regulation was confirmed in PCR (the up-regulated ones mentioned by the authors) were also detected in the re-analysis.

Besides, Nielsen *et al.* (2009) had to focus on miRNAs they had found to be differentially expressed. Only for those with a large enough fold change they regarded the gene sets. Requiring that a miRNA needs to be significantly differentially expressed leads to a quite strict procedure, however. miRNAs which are just slightly differential or whose fold change is rather low may have the strongest imaginable impact on their targets - one would still miss them. Consequently, it is not surprising that more miRNAs were detected in the re-analysis, also such which would not have been detected by miRNA-wise testing without a fold change threshold. Depending on the gene set test applied this was a reasonable number of miRNAs (for enrichment and rotation tests) or nearly all of them (for the global tests). The rotation tests, as the most conservative ones, detected the smallest number of miRNAs that had not been previously mentioned by Nielsen *et al.*

(2009). This may be a general property. However, even though it does not seem likely, it is also possible that they were limited to show low p-values because of the limited amount of random rotations applied.

Interestingly, unlike in the simulations, here it was primarily the rotation tests which were able not to detect miRNAs one may not wish to find. One of these was especially miR-290, which was down-regulated and had been shown by Nielsen *et al.* (2009) to have a target set differential in the same direction. This result was confirmed by the fact that it was also not detected by the Wilcoxon-based procedure. Additionally, miR-218 shall be named, for which the gene set displayed the same non-inverse regulation and which was only not detected by the rotation tests. Their comparably conservative behaviour regarding the two miRNAs might be caused by their compensation for inter-gene correlation. However, they also missed miRNAs that may have been of interest or for which at least no obvious reason not to detect them existed, as for example miR-124a which was up-regulated and had had a significantly down-regulated subset of target-mRNAs according to Nielsen *et al.* (2009).

4.3   Conclusion and Outlook

To conclude, the idea of joining miRNA and mRNA expression as well as miRNA target annotation data to create a procedure for two-groups comparisons seems to work. Generally, it performed well both on synthetic and on real data. It was implemented in the R statistical programming language (R Development Core Team, 2010) under the package name 'miRtest'. The can be found on the CD attached. In its newest version, it is also downloadable from CRAN (http://cran.r-project.org/ R Development Core Team, 2010).

As mentioned before the idea can be further investigated when it comes to the tests' behaviour under different simulation settings than the ones applied above. For the problem of correlating genes one might try an approach that includes permutation of the group membership of each microarray and an enrichment test. Even though this would solve the problem of dependent statistics, it would be time consuming. Furthermore, to show low p- or even low q-values it would depend on much larger numbers of microarray replicates

than usually exist in a typical data set.

This work concentrated on two-group testing. However, the test can easily be gener-alised to any linear model, as both the methodology of Smyth (2004), as well as of the gene set tests allow for more groups or continuous variables. An example of the latter would be quantitative measures such as survival times or body mass index associated with each microarray replicate. This is already implemented in 'miRtest' to a certain degree.

Another point that was not dealt with here are covariables. They include additional information one has on the data set, such as age or gender of the animals sacrificed for the microarray experiments. To neglect their influence means to lose power. They are also partly implemented in the 'miRtest' package already, but only for enrichment and rotation tests. Their full incorporation and application on synthetic data might be of interest for future work as well.

Apart from that matched data is a major issue. Many experiments are designed in such a way that miRNA and mRNA samples come from the same origin, be it two samples of the same cancerous tissue or the same animal sacrificed. However, both combination methodologies applied, Fisher's and Stouffer's method, depend strictly speaking on two independent p-values. If this is not the case the p-values they generate can become skewed. The solution for such a problem was presented above numerous times: resampling. Consequently, an idea would be to, given matched data, jointly permute the columns of the expression matrices of miRNAs and mRNAs. If the number of possible permutations does not suffice, one could think of joint rotating. This would be more sophisticated, however, as the assumption of a joint multivariate normal distribution of both expression matrices would have to be made. That would have at least implications on the way data needs to be normalised beforehand.

A final suggestion for improvement of the algorithm shall be made concerning the miRNA-mRNA network itself. So far, the graph is used for testing for significance but hardly integrated into multiple testing correction. Approaches to do so exist, however. See Goeman and Mansmann (2008) or Bauer *et al.* (2010) for examples on the graph of gene ontology. To include such approaches into miRNA testing might be another goal of

future work.

References

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., Dreyfuss, G., Eddy, S. R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., and Tuschl, T. (2003). A uniform system for microRNA annotation. *RNA (New York, N.Y.)*, **9**(3), 277–279. PMID: 12592000.

Bauer, S., Gagneur, J., and Robinson, P. N. (2010). GOing bayesian: model-based gene set analysis of genome-scale data. *Nucleic Acids Research*, **38**(11).

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**(1).

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, **29**(4).

Bland, J. M. and Altman, D. G. (1995). Multiple significance tests: the bonferroni method. *BMJ*, **310**(6973), 170.

Bohnsack, M. T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA (New York, N.Y.)*, **10**(2), 185–191. PMID: 14730017.

Bolstad, B., Irizarry, R., strand, M., and Speed, T. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**(2), 185 –193.

Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., Holloway, E., Kapushesky, M., Kemmeren, P., Lara, G. G., Oezcimen, A., Rocca-Serra, P., and Sansone, S. (2003). ArrayExpress–a public repository for microarray gene expression data at the EBI. *Nucleic Acids Research*, **31**(1).

Brennecke, J., Stark, A., Russell, R. B., and Cohen, S. M. (2005). Principles of MicroRNATarget recognition. *PLoS Biol*, **3**(3), e85.

Brunner, E. (2009). Repeated measures under non-sphericity. *Proceedings of the National Academy of Sciences of the United States of America*, pages 605–609.

CAI, X., HAGEDORN, C. H., and CULLEN, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, **10**(12), 1957 –1966.

Cannell, I. G., Kong, Y. W., and Bushell, M. (2008). How do microRNAs regulate gene expression? *Biochemical Society Transactions*, **36**(Pt 6), 1224–1231. PMID: 19021530.

Chang, T. and Mendell, J. T. (2007). microRNAs in vertebrate physiology and human disease. *Annual Review of Genomics and Human Genetics*, **8**, 215–239. PMID: 17506656.

Dudoit, S., Shaffer, J., and Boldrick, J. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science*, **18**(1), 71–103.

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

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in drosophila. *Genome Biology*, **5**(1), R1. PMID: 14709173.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes & Development*, **21**(20), 2558 –2570.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2009). Deadenylation is a widespread effect of miRNA regulation. *RNA (New York, N.Y.)*, **15**(1), 21–32. PMID: 19029310.

Fisher, R. A. (1950). Statistical methods for research workers. (11th ed.), 99–101.

Gatti, D., Barry, W., Nobel, A., Rusyn, I., and Wright, F. (2010). Heading down the wrong pathway: on the influence of correlation within gene sets. *BMC Genomics*, **11**(1), 574.

Goeman, J. J. and Bühlmann, P. (2007). Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics (Oxford, England)*, **23**(8).

Goeman, J. J. and Mansmann, U. (2008). Multiple testing on the directed acyclic graph of gene ontology. *Bioinformatics*, **24**(4), 537 –544.

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

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, **34**(Database issue).

Griffiths-Jones, S., Saini, H. K., van Dongen, S., and Enright, A. J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Research*, **36**(Database issue), D154–158. PMID: 17991681.

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Molecular Cell*, **27**(1), 91–105.

Hummel, M., Meister, R., and Mansmann, U. (2008). GlobalANCOVA: exploration and assessment of gene group effects. *Bioinformatics*, **24**(1), 78 –85.

Jayaswal, V., Lutherborrow, M., Ma, D., and Yang, Y. H. (2009). Identification of microRNAs with regulatory potential using a matched microRNA-mRNA time-course data. *Nucl. Acids Res.*, **37**(8), e60.

Lee, Y., Kim, M., Han, J., Yeom, K., Lee, S., Baek, S. H., and Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*, **23**(20), 4051–4060.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**(5689), 1437–1441.

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R., and Golub, T. R. (2005). MicroRNA expression profiles classify human cancers. *Nature*, **435**(7043), 834–838.

Mansmann, U. and Meister, R. (2005). Testing differential gene expression in functional groups. goeman's global test versus an ANCOVA approach. *Methods of Information in Medicine*, **44**(3).

Marot, G. and Mayer, C. (2009). Sequential analysis for microarray data based on sensitivity and Meta-Analysis. *Statistical Applications in Genetics and Molecular Biology*, **8**(1).

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC, 2 edition.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell*, **15**(2), 185–197. PMID: 15260970.

Nielsen, J. A., Lau, P., Maric, D., Barker, J. L., and Hudson, L. D. (2009). Integrating microRNA and mRNA expression profiles of neuronal progenitors to identify regulatory networks underlying the onset of cortical neurogenesis. *BMC Neuroscience*, **10**.

Nykter, M., Aho, T., Ahdesmaki, M., Ruusuvuori, P., Lehmussola, A., and Yli-Harja, O. (2006). Simulation of microarray data with realistic characteristics. *BMC Bioinformatics*, **7**(1), 349.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L., and Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome Research*, **14**(10a), 1902 –1910.

Seitz, H. and Zamore, P. D. (2006). Rethinking the microprocessor. *Cell*, **125**(5), 827–829. PMID: 16751089.

Sethupathy, P., Megraw, M., and Hatzigeorgiou, A. G. (2006). A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat Meth*, **3**(11), 881–886.

Shi, J. (May 2007). Gene set enrichment analysis (gsea) for interpreting gene expression profiles. *Current Bioinformatics*, **2**, 133–137(5).

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

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., and Williams, R. M. (1949). Adjustment during army life. In *The American Soldier*, volume Vol. 1. Princeton University Press, Princeton.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(43).

Volinia, S., Calin, G. A., Liu, C., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., Prueitt, R. L., Yanaihara, N., Lanza, G., Scarpa, A., Vecchione, A., Negrini, M., Harris, C. C., and Croce, C. M. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(7), 2257 –2261.

Wei, C. (2010). *Basiswissen Medizinische Statistik*. Springer, Berlin, 5., berarbeitete auflage. edition.

Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M., Visvader, J. E., and Smyth, G. K. (2010). ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics (Oxford, England)*, **26**(17).

Wu, L., Fan, J., and Belasco, J. G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(11), 4034–4039. PMID: 16495412.

Acknowledgement

This is the Acknowledgement.

Appendix

Media Content

On the Compact Disc (CD) attached this work can be found in .pdf format as well as the package 'miRtest' written in the R statistical programming language R Development Core Team (2010) that implements the algorithms described above. Furthermore, as they are too many to be put in the appendix, the Figures which show FDR and APR of all simulations performed and a table with the q-values of all the miRNAs of the data example can be found on the disc as well.

The CD's directory structure is as follows: The main directory contains

- the Thesis as a file named 'thesis.pdf',

- the subdirectory 'dataExample' which contains the miRNA lists of the data example in .pdf and .csv format,

- a subdirectory named 'figures' with the plots of the simulations. It has four subdirectories corresponding to the four simulations performed, namely

    'noOverlapWeak' (no overlaps, weak gene set effect),

    'noOverlapStrong' (no overlaps, strong gene set effect),

    'OverlapWeak' (overlaps, weak gene set effect) and

    'OverlapStrong' (overlaps, strong gene set effect).

The plots in the subdirectories of 'figures' are in .pdf file format. Each file contains one plot. The name of each file is structured as follows:

rate.test.cov.pdf, which stand for

rate	FDR or APR
test	Procedure (dep. on which Gene Set Test)
cov	Covariance Structure of the Synthetic Data tested

Statutory Declaration

I hereby assure that I have composed the present Thesis entitled

Evaluating miRNA and target-mRNA Expression Profiles

independently and have used no other appliances than indicated. Parts being gathered from other works according to wording or meaning I have indicated in every single case by declaration of source.

I hereby state furthermore, that I have produced my works according to the principles of good scientific practise in compliance with the valid 'Richtlinien der Georg-August-Universität zur Sicherung guter wissenschaftlicher Praxis'.

Place, Date: _____

Signature: _____

(Stephan Artmann)