

Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons

Bettina Harr* and Christian Schlötterer¹

Institut für Genetik, Universität Köln, Zùlpicherstrasse 47, 50674 Köln, Germany and ¹Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, Veterinärplatz 1, 1210 Wien, Austria

Received September 25, 2005; Revised November 9, 2005; Accepted January 3, 2006

ABSTRACT

Oligonucleotide microarrays are an informative tool to elucidate gene regulatory networks. In order for gene expression levels to be comparable across microarrays, normalization procedures have to be invoked. A large number of methods have been described to correct for systematic biases in microarray experiments. The performance of these methods has been tested only to a limited extent. Here, we evaluate two different types of microarray analyses: (i) the same gene in replicate samples and (ii) different, but co-expressed genes in the same sample. The reliability of the latter analysis needs to be determined for the analysis of regulatory networks and our report is the first attempt to evaluate for the accuracy of different microarray normalization methods in this respect. Consistent with previous results we observed a large effect of the normalization method on the outcome of the expression analyses. Our analyses indicate that different normalization methods should be performed depending on whether a study is aiming to detect differential gene expression between independent samples or whether co-expressed genes should be identified. We make recommendations about the most appropriate method to use.

INTRODUCTION

Complex biological processes require the interaction of many different genes. In order to understand the role of individual genes or gene products in a biological process, knowledge of genome-wide gene expression patterns is required. Microarrays measure expression levels of thousands of genes in a

single experiment, thus providing a powerful tool to elucidate gene regulatory networks (1). One version of microarrays, high-density oligonucleotide microarrays (Affymetrix chips), uses oligonucleotides with length of 25 bp to assay transcription levels at individual genes. On the *Escherichia coli* Affymetrix microarray each gene is represented by 14–20 pairs of oligonucleotides. Each pair of oligos consists of a perfect match (PM) and a mismatch probe (MM), the latter being identical to the former with the exception of a single mismatch in the central position of the oligo. The purpose of MM oligos on Affymetrix chips is to correct for non-specific binding of the mRNA. In order to obtain expression levels of genes, chips are hybridized with fluorescently labeled RNA and scanned. Fluorescent intensities of each oligo pair are then algorithmically combined to yield a single expression value for each gene.

In microarray experiments there are many sources of systematic variation. These might be linked to differences in probe labeling efficiency, RNA concentration or hybridization efficiency. To correct for this, numerous ‘normalization’ methods have been proposed. Among the most commonly used methods are the Affymetrix Microarray Suite 5 method (*mas5.0*) [<http://www.affymetrix.com/support/technical/whitepapers.affx>, (2)], the perfect match only model of Li and Wong (3,4), the Robust Multi-array Analysis with [*gcrma*, (5)] and without [*rma*, (6)] correction for GC content of the oligo. Previous work compared different normalization methods and found a profound impact on which one was used for detection of differentially expressed genes (7,8).

One problem with testing normalization methods is that no ‘golden standard’ exists to which expression values can be compared. In an attempt to evaluate the performance of the methods, ‘spike-in’ experiments have been conducted in which known concentrations of mRNA are added to the hybridization cocktail (9). The performance of a method is then judged by how well it identifies different concentrations of mRNA. From these studies (9) concluded the *rma* method is superior in terms of sensitivity and specificity (i.e. the true

*To whom correspondence should be addressed. Tel: +49 221 470 2324; Fax: +49 221 470 5975; Email: harrb@uni-koeln.de

and false detection rate) to *mas5.0* and the method of Li and Wong. One drawback of spike-in experiments is that they themselves include sources of systematic variation (e.g. in hybridization efficiencies). It is not clear how the evaluation of different methods would be affected by such systematic differences. One alternative assay that has been proposed is to compare transcription levels between males and females at a set of Y-chromosome linked genes, thus providing a true biological control (10). In this study the performance of a normalization method is assayed by recording how many differentially expressed Y-chromosome genes are recovered in an experiment involving male and female samples. However, the power of this test is quite small given that out of 45 Y-linked genes, 11 could be identified by one method and 9 were identified by the other method.

Here, we propose an alternative strategy to evaluate normalization methods making use of the fact that bacterial genes are organized in operons. In an operon two or more adjacent genes are co-transcribed into a single mRNA. Thus, genes located in a given operon are expected to be highly correlated in their expression level. This fact provides a basis for a test of which normalization method would best predict this correlation. We assay correlation coefficients in gene expression among members of known operons in *E.coli*. Since a single transcript yields expression levels of several genes this method is not expected to be affected by systematic biases such as the ones mentioned above.

MATERIALS AND METHODS

Microarray data

The data will be described in detail elsewhere (B. Harr and C. Schlötterer, manuscript in preparation). In brief, four different *E.coli* variants were investigated using Affymetrix oligonucleotide microarrays: the DH5 α strain containing an F plasmid ([F- Φ 80d*lacZ* Δ M15 Δ (*lacZYA-argF*) U169 *end A1 recA 1 hsdR17*(r_k - m_k +) *deoR thi-1 supE44 λ -gyrA96 relA1*]), DH5 α without F-plasmid, the MG1655 strain (12) containing an F-plasmid and MG1655 without an F-plasmid. Replicate cultures for each genotype were inoculated from single colonies in 5 ml Luria-Bertani (LB) medium (0.2 mg/ml Ampicillin) and grown overnight at 37°C. Of each overnight culture 500 μ l was used to inoculate 50 ml fresh LB medium. These cultures were grown at 37°C and cells were harvested in early log phase corresponding to an OD₆₀₀ of 0.4. Total RNA was extracted using the MasterPure™ RNA Purification Kit obtained from Epicenter (Catalog no. MCR85102) from 1 ml early log phase culture, fragmented, 3' end-labeled and hybridized to an Affymetrix *E.coli* Antisense Array. Patterns of hybridization were detected with an Affymetrix scanner. Every *E.coli* open reading frame (11) is assayed on the Affymetrix Chip by a set of PM and MM probe pairs.

Analysis

Raw signal intensities for each probe set as they are contained in the *CEL* files were analyzed using a series of methods implemented in the software package Bioconductor (<http://bioconductor.org>). Altogether we analyzed the data using 54 methods, which consisted of various combinations of

4 background correction algorithms (*none*, *mas*, *rma* and *gcrma*), 3 normalization algorithms (*constant*, *quantiles* and *invariantset*), 2 perfect match correction algorithms (*pmonly* and *mas*) and 3 summary algorithms (*mas*, *liwong* and *medianpolish*). The methods and references to the methods are described in detail in <http://www.bioconductor.org> (12). The choice of the algorithms was motivated by the fact that some combinations of these should result in commonly used summaries of Affymetrix microarray data (i.e. the *mas 5.0* normalization method, the *rma* and *gcrma* normalization method and the *Li-Wong* normalization method). The *rma* and *gcrma* background correction method adjusts only perfect match signals. Thus, it can only be used in conjunction with the *pmonly* PM correction method. For all analysis we used only genes expressed in replicate samples of one strain (i.e. samples had a 'Present' Affymetrix detection call or an Affymetrix MAS5.0 intensity of >200).

Comparison of correlation coefficients within and between replicates

For each of the four *E.coli* strains 16 Spearman correlation coefficients were calculated. The first coefficients were calculated between data from a single slide analyzed with two different normalization methods (Figure 1A). Here we considered only a subset of four different analysis methods, namely those that correspond to the standard Affymetrix method *mas5.0*, *rma*, *gcrma* and *Li-Wong*. Twelve such *between method* correlation coefficients were calculated for each *E.coli* strain (i.e. correlation coefficients between all possible pairwise combinations of two normalization methods applied to the same slide). The remaining four correlation coefficients (*within method* correlations, Figure 1B) describe the correlation between replicate experiments of one strain analyzed with a single method (either *mas5.0*, *rma*, *gcrma* or *Li-Wong*). The difference between the groups of *between method* and *within method* correlation coefficients were compared by a Mann-Whitney *U*-test.

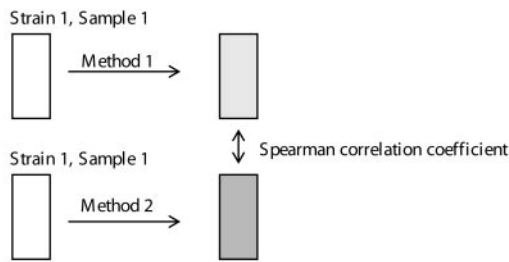
Operon correlation

All known operons with at least two member genes (345 as of October 2005) were downloaded from RegulonDB (13) using the GETools webpage (http://www.cifn.unam.mx/Computational_Genomics/regulondb/) (14). About 42% of operons had only two member genes. If operons had more than two members, we randomly selected two genes from each operon. Only those operons where both members were expressed in the particular strain were used for further analysis. Out of the 345 known operons 208 were expressed in DH5 α , 205 in DH5 α F, 213 in MG1655 and 189 in MG1655F under our experimental conditions. For each sample (i.e. two samples per *E.coli* strain), each operon and each of the 54 analysis methods, we calculated the Spearman rank correlation coefficients between signal intensities of operon gene 1 against operon gene 2 (Figure 1C).

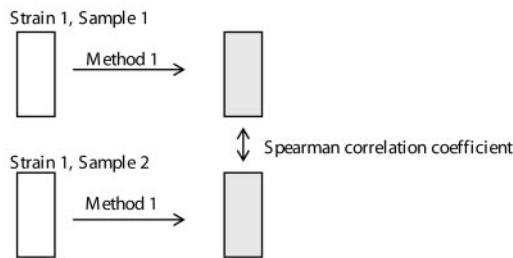
Correlations between replicates

In analogy to the operon correlation we tested the performance of all 54 analysis methods by calculating a Spearman rank correlation coefficient between replicate samples within one *E.coli* strain (Figure 1B).

A. "Between method" correlations



B. "Within method" correlations



C. Correlation among operon members

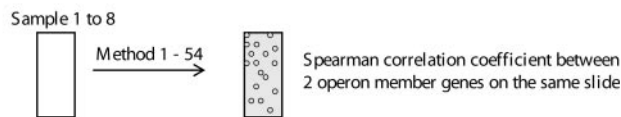


Figure 1. Graphical representation of the analyses performed in this study. (A) Each sample is analyzed with four different normalization methods and correlation coefficients are calculated for pairwise combinations of normalization methods within one sample. (B) Replicate samples within each strain are analyzed with four different normalization methods and correlation coefficients are calculated between the replicate samples analyzed with a single method. (C) Each of the eight samples (i.e. four *E.coli* strains \times two hybridizations per strain) was analyzed independently under 54 different normalization methods. For each method the correlation in expression level between two operon member genes is calculated.

RESULTS

Following the methods outlined in references (15–17) we used correlation coefficients to assay the performance of different methods. Both parametric and non-parametric correlations have been used. The Spearman correlation uses ranks rather than raw expression levels which makes it less sensitive to extreme values in the data. In this study we present non-parametric Spearman rank correlation coefficients as a robust method with which to compare results from replicating an array and those from analyzing raw data from a single experiment employing different normalization approaches. In this first analysis only four normalization methods were considered (*mas5.0*, *rma*, *gcrma* or *Li-Wong*). As shown in Table 1 coefficients that describe the correlation between a single array normalized according to different normalization methods (*between method*) are significantly worse than those calculated for replicate arrays using a single method (*within method*,

$P < 0.0001$, Mann–Whitney *U*-test). For the correlation coefficients calculated using *within method* we found that the *mas5.0* methods performs significantly worse than the *gcrma* ($P = 0.0304$, Mann–Whitney *U*-test) and *Li-Wong* ($P = 0.0313$, Mann–Whitney *U*-test) method. No significant difference could be detected between other pairwise comparisons of normalization methods. Thus, consistent with previous results (8), normalization methods have a profound effect which goes beyond the variance that can be observed across replicate arrays.

To investigate the correlation coefficients as a function of signal intensity we separated the full dataset in six different signal intensity classes with each of the six classes containing an equal number of genes. For all signal intensity classes we found the same relationships as described above, i.e. the *between method* correlation coefficients are significantly lower than the *within method* coefficients and *gcrma* and *Li-Wong* perform best in the *within method* analysis (data not shown). Moreover, the correlation coefficients for the different intensity classes are highly significantly positively correlated with each other (e.g. Spearman correlation coefficient between correlation coefficients in the lowest and highest intensity class, $P \leq 0.001$). Thus signal intensity does not seem to have a profound impact on our results.

There are four steps that need to be carried out during a microarray normalization procedure: background correction (to adjust for hybridization effects that are not associated with the interaction of probe with target DNA), normalization (adjust chips to a common baseline so that they are comparable among each other), perfect/mismatch correction (adjust for non-specific hybridization of target to probe) and summarization (summarize individual oligo-pairs to yield a single expression value per gene). The Bioconductor package provides the opportunity to combine these four normalization steps to yield a single expression value for each gene.

Correlation among operon member genes

We tested 54 different analysis methods corresponding to combinations of normalization steps by calculating correlation coefficients among operon members. These values are reported in Table 2 together with a mean correlation coefficient over all four strains and replicate samples within a given analysis method. Table 2 shows a ranked list of the correlation coefficients. Interestingly, the method we found that works best does not correspond to either of the commonly used normalization methods *mas5.0*, *rma*, *gcrma* or *Li-Wong*. Instead it employs a combination of the *mas5.0* and the *Li-Wong* methods (i.e. background correction and perfect/mismatch correction according to *mas5.0*, and normalize and summary method according to the *Li-Wong* method). Among the commonly employed methods, *Li-Wong* performs best and *rma* performs worst, *mas5.0* and *gcrma* perform equally well and are intermediate between *Li-Wong* and *rma*.

Correlation between replicate samples

Table 3 shows the ranked list of correlation coefficients (individual values for each strain and averages across *E.coli* strains) obtained from replicating an array for the 54 different normalization methods. In contrast to the correlation among operon member genes the *gcrma* and *rma* methods perform

Table 1. Spearman correlation coefficients calculated on two levels

| Method | Sample | | DH α | DH α F | MG1655 | MG1655F |
|-------------------------------------|--------------------------|-----------------------|-------------|---------------|--------|---------|
| <i>gcrma</i> | Sample 1 versus Sample 2 | <i>within method</i> | 0.98 | 0.99 | 0.99 | 0.97 |
| <i>rma</i> | Sample 1 versus Sample 2 | <i>within method</i> | 0.96 | 0.98 | 0.98 | 0.96 |
| <i>Li-Wong</i> | Sample 1 versus Sample 2 | <i>within method</i> | 0.95 | 0.97 | 0.98 | 0.95 |
| <i>mas5.0</i> | Sample 1 versus Sample 2 | <i>within method</i> | 0.94 | 0.96 | 0.97 | 0.93 |
| <i>rma</i> versus <i>mas5.0</i> | Sample 1 | <i>between method</i> | 0.94 | 0.95 | 0.96 | 0.91 |
| <i>rma</i> versus <i>Li-Wong</i> | Sample 1 | <i>between method</i> | 0.89 | 0.91 | 0.92 | 0.87 |
| <i>mas5.0</i> versus <i>Li-Wong</i> | Sample 1 | <i>between method</i> | 0.87 | 0.90 | 0.92 | 0.83 |
| <i>mas5.0</i> versus <i>gcrma</i> | Sample 1 | <i>between method</i> | 0.81 | 0.86 | 0.85 | 0.79 |
| <i>rma</i> versus <i>gcrma</i> | Sample 1 | <i>between method</i> | 0.74 | 0.78 | 0.79 | 0.68 |
| <i>Li-Wong</i> versus <i>gcrma</i> | Sample 1 | <i>between method</i> | 0.72 | 0.78 | 0.78 | 0.69 |
| <i>rma</i> versus <i>mas5.0</i> | Sample 2 | <i>between method</i> | 0.95 | 0.94 | 0.96 | 0.95 |
| <i>rma</i> versus <i>Li-Wong</i> | Sample 2 | <i>between method</i> | 0.92 | 0.89 | 0.93 | 0.90 |
| <i>mas5.0</i> versus <i>Li-Wong</i> | Sample 2 | <i>between method</i> | 0.91 | 0.88 | 0.93 | 0.89 |
| <i>mas5.0</i> versus <i>gcrma</i> | Sample 2 | <i>between method</i> | 0.87 | 0.82 | 0.86 | 0.86 |
| <i>rma</i> versus <i>gcrma</i> | Sample 2 | <i>between method</i> | 0.79 | 0.73 | 0.80 | 0.78 |
| <i>Li-Wong</i> versus <i>gcrma</i> | Sample 2 | <i>between method</i> | 0.78 | 0.73 | 0.80 | 0.77 |

I. *Within method*, between two replicates analyzed with the same normalization method; II. *Between method*, calculated within a single replicate analyzed for different pairwise combinations of microarray normalization methods. Note that only four different normalization schemes were used (*mas5.0*, *rma*, *gcrma* and *Li-Wong*, see text for references).

much better than the two other commonly used normalization methods (*Li-Wong* and *mas5.0*).

To determine the basis underlying this difference in correlations between replicates and correlations between operon members within a single microarray we analyzed each normalization step individually in a series of non-parametric tests. Figure 2A shows the case for the correlations among operon member genes. The background correction and normalizing methods did not have a profound impact on the correlation coefficient. In contrast, the incorporation of mismatch oligo signals does seem to be performing better than only incorporating perfect match signals ($P = 0.0113$, Mann-Whitney U -test). A strong influence of the summary method was detected. The *Li-Wong* summary method performed significantly better than the *mas5.0* and *rma/gcrma* version of the summary method ($P < 0.0001$). The *rma/gcrma* version of the summary method (i.e. *medianpolish*) was significantly worse than *mas5.0* ($P < 0.0114$).

Similar to correlations among operon member genes within a single experiment, the correlation among replicates was little affected by the background correction and normalize method, but strongly affected by the perfect match/mismatch correction method and by the summarize method (Figure 2B). However, the latter two effects are in opposite direction than in the operon correlation case with the perfect match correction algorithm *pmonly* and the summary algorithm *medianpolish* are associated with the strongest correlations between replicate arrays. Both are implemented in the normalizing method *rma* and *gcrma* and therefore explain the superiority of this method in describing the correlation between replicates.

DISCUSSION

Previously, it has been found that genes located in operons are strongly correlated in expression values. Using Hidden Markov Models on gene expression data, Tjaden *et al.* (18) identified operon elements, which corresponded to known operons in *E.coli* with high confidence. Here we use this result

to test different normalization methods. A potential limitation is that operon member genes might have low correlations if transcription is driven by additional internal promoters. Moreover, transcription units may overlap each other and regulatory elements could overlap transcription units or other regulatory elements (19). However, these processes should not have a systematic basis and are unlikely to affect our results.

We analyzed each of the four steps in the normalization procedure separately and, within one step, determined the algorithm associated with the highest correlation coefficient between expression levels. This procedure was performed twice. First, we determined correlations between co-expressed genes (i.e. genes transcribed in the same operon) within a single microarray. The second analysis concerned correlations in expression levels between identical genes on two different replicate arrays.

In the case of analysis of co-expression, we found a strong influence of summary method with the *Li-Wong* (3) method performing significantly best. The *Li-Wong* method was originally motivated by the observation that the information on expression level provided by the different probes for the same gene are highly variable, even if the intensity information from the mismatch oligo is taken into consideration. To account for these probe-specific effects a Model Based Expression Index was introduced. The algorithm iteratively fits a model to the probe set data from multiple microarrays. In this way, cross hybridizing probes, arrays with image contaminations at certain probe sets and single outliers can be excluded and replaced by fitted values. Thus, in order to identify co-expression among genes it is most important that the unique responsiveness of each probe is taken into account.

It is desirable to achieve high reproducibility across replicate chips if we wish to detect differentially expressed genes between different experiments. We found that among the four steps in the normalization procedure only two had a significant effect on correlations in expression levels between replicate microarrays. Specifically, the perfect match only version of the perfect match correction method and the *rma/gcrma* version of the summarize method performed best.

Table 2. Spearman correlation coefficients between operon member genes

| MG1 | MG2 | MGF1 | MGF2 | DH1 | DH2 | DHF1 | DHF2 | Average | Method | Normalize | pm correct | Summary | Common name |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|---------------|--------------|---------------------|---------------|---------------------|----------------|
| <i>r</i> | <i>r</i> | <i>r</i> | <i>r</i> | <i>r</i> | <i>r</i> | <i>r</i> | <i>r</i> | | Background | | | | |
| 0.69 | 0.73 | 0.71 | 0.67 | 0.71 | 0.72 | 0.63 | 0.68 | 0.6921 | mas | invariantset | mas | liwong | |
| 0.68 | 0.73 | 0.71 | 0.67 | 0.71 | 0.73 | 0.62 | 0.67 | 0.6877 | none | quantiles | mas | liwong | |
| 0.68 | 0.73 | 0.71 | 0.67 | 0.71 | 0.72 | 0.62 | 0.67 | 0.6877 | mas | quantiles | mas | liwong | |
| 0.68 | 0.73 | 0.71 | 0.66 | 0.70 | 0.72 | 0.62 | 0.66 | 0.6867 | none | invariantset | mas | liwong | |
| 0.66 | 0.70 | 0.69 | 0.64 | 0.70 | 0.72 | 0.60 | 0.65 | 0.6700 | mas | quantiles | pmonly | liwong | |
| 0.67 | 0.71 | 0.69 | 0.65 | 0.69 | 0.71 | 0.60 | 0.64 | 0.6697 | none | constant | mas | liwong | |
| 0.66 | 0.70 | 0.69 | 0.64 | 0.70 | 0.72 | 0.60 | 0.65 | 0.6694 | rma | quantiles | pmonly | liwong | |
| 0.65 | 0.70 | 0.68 | 0.65 | 0.69 | 0.70 | 0.60 | 0.64 | 0.6651 | mas | constant | mas | liwong | |
| 0.65 | 0.69 | 0.68 | 0.63 | 0.70 | 0.72 | 0.60 | 0.64 | 0.6632 | rma | invariantset | pmonly | liwong | |
| 0.65 | 0.69 | 0.68 | 0.63 | 0.69 | 0.71 | 0.59 | 0.64 | 0.6610 | none | quantiles | pmonly | liwong | |
| 0.65 | 0.69 | 0.67 | 0.62 | 0.69 | 0.71 | 0.59 | 0.63 | 0.6574 | mas | invariantset | pmonly | liwong | |
| 0.63 | 0.68 | 0.65 | 0.61 | 0.68 | 0.70 | 0.58 | 0.63 | 0.6456 | none | constant | pmonly | liwong | |
| 0.64 | 0.68 | 0.66 | 0.60 | 0.69 | 0.70 | 0.56 | 0.62 | 0.6443 | none | invariantset | pmonly | liwong | Li-Wong |
| 0.63 | 0.68 | 0.65 | 0.60 | 0.68 | 0.70 | 0.57 | 0.63 | 0.6415 | rma | constant | pmonly | liwong | |
| 0.63 | 0.67 | 0.65 | 0.59 | 0.67 | 0.70 | 0.57 | 0.62 | 0.6397 | mas | constant | pmonly | liwong | |
| 0.65 | 0.67 | 0.64 | 0.61 | 0.66 | 0.68 | 0.60 | 0.60 | 0.6379 | gcrma | invariantset | pmonly | liwong | |
| 0.65 | 0.68 | 0.65 | 0.59 | 0.65 | 0.67 | 0.60 | 0.60 | 0.6362 | gcrma | constant | pmonly | mas | |
| 0.64 | 0.68 | 0.65 | 0.59 | 0.65 | 0.67 | 0.60 | 0.60 | 0.6352 | gcrma | quantiles | pmonly | mas | |
| 0.64 | 0.68 | 0.65 | 0.59 | 0.65 | 0.67 | 0.60 | 0.60 | 0.6352 | gcrma | invariantset | pmonly | mas | |
| 0.63 | 0.68 | 0.63 | 0.59 | 0.65 | 0.68 | 0.59 | 0.61 | 0.6336 | none | invariantset | mas | mas | |
| 0.62 | 0.68 | 0.63 | 0.59 | 0.66 | 0.69 | 0.59 | 0.61 | 0.6335 | none | quantiles | mas | mas | |
| 0.65 | 0.67 | 0.63 | 0.61 | 0.66 | 0.67 | 0.58 | 0.59 | 0.6316 | gcrma | quantiles | pmonly | liwong | |
| 0.62 | 0.68 | 0.63 | 0.59 | 0.65 | 0.68 | 0.57 | 0.61 | 0.6302 | none | constant | mas | mas | |
| 0.62 | 0.68 | 0.63 | 0.58 | 0.66 | 0.68 | 0.58 | 0.61 | 0.6301 | mas | invariantset | mas | mas | |
| 0.62 | 0.68 | 0.63 | 0.58 | 0.66 | 0.68 | 0.58 | 0.61 | 0.6295 | mas | quantiles | mas | mas | |
| 0.62 | 0.68 | 0.63 | 0.58 | 0.66 | 0.68 | 0.56 | 0.61 | 0.6265 | mas | constant | mas | mas | mas5.0 |
| 0.63 | 0.66 | 0.63 | 0.59 | 0.64 | 0.67 | 0.58 | 0.60 | 0.6256 | none | constant | mas | medianpolish | |
| 0.65 | 0.66 | 0.64 | 0.60 | 0.63 | 0.64 | 0.59 | 0.58 | 0.6249 | gcrma | quantiles | pmonly | medianpolish | gcrma |
| 0.65 | 0.66 | 0.64 | 0.60 | 0.62 | 0.65 | 0.59 | 0.58 | 0.6243 | gcrma | invariantset | pmonly | medianpolish | |
| 0.63 | 0.66 | 0.63 | 0.59 | 0.64 | 0.67 | 0.58 | 0.59 | 0.6239 | mas | constant | mas | medianpolish | |
| 0.65 | 0.66 | 0.64 | 0.59 | 0.62 | 0.65 | 0.58 | 0.59 | 0.6228 | gcrma | constant | pmonly | medianpolish | |
| 0.62 | 0.66 | 0.63 | 0.59 | 0.64 | 0.67 | 0.58 | 0.60 | 0.6226 | none | quantiles | mas | medianpolish | |
| 0.64 | 0.65 | 0.63 | 0.60 | 0.64 | 0.66 | 0.57 | 0.58 | 0.6224 | gcrma | constant | pmonly | liwong | |
| 0.62 | 0.66 | 0.62 | 0.59 | 0.64 | 0.66 | 0.58 | 0.59 | 0.6208 | mas | quantiles | mas | medianpolish | |
| 0.62 | 0.66 | 0.62 | 0.58 | 0.64 | 0.66 | 0.57 | 0.59 | 0.6183 | none | invariantset | mas | medianpolish | |
| 0.62 | 0.66 | 0.62 | 0.58 | 0.63 | 0.66 | 0.57 | 0.59 | 0.6181 | mas | invariantset | mas | medianpolish | |
| 0.61 | 0.66 | 0.62 | 0.57 | 0.64 | 0.68 | 0.55 | 0.60 | 0.6167 | none | invariantset | pmonly | mas | |
| 0.61 | 0.66 | 0.62 | 0.57 | 0.65 | 0.68 | 0.55 | 0.60 | 0.6166 | none | quantiles | pmonly | mas | |
| 0.61 | 0.66 | 0.62 | 0.57 | 0.64 | 0.68 | 0.55 | 0.60 | 0.6158 | none | constant | pmonly | mas | |
| 0.61 | 0.66 | 0.62 | 0.57 | 0.64 | 0.68 | 0.55 | 0.60 | 0.6157 | mas | constant | pmonly | mas | |
| 0.61 | 0.67 | 0.62 | 0.57 | 0.64 | 0.68 | 0.55 | 0.60 | 0.6156 | mas | quantiles | pmonly | mas | |
| 0.61 | 0.67 | 0.61 | 0.57 | 0.64 | 0.68 | 0.55 | 0.60 | 0.6150 | mas | invariantset | pmonly | mas | |
| 0.61 | 0.67 | 0.61 | 0.57 | 0.64 | 0.68 | 0.55 | 0.59 | 0.6150 | rma | quantiles | pmonly | mas | |
| 0.61 | 0.66 | 0.61 | 0.57 | 0.64 | 0.68 | 0.54 | 0.59 | 0.6145 | rma | constant | pmonly | mas | |
| 0.62 | 0.65 | 0.62 | 0.57 | 0.63 | 0.67 | 0.57 | 0.59 | 0.6145 | mas | constant | pmonly | medianpolish | |
| 0.61 | 0.66 | 0.61 | 0.57 | 0.64 | 0.68 | 0.54 | 0.60 | 0.6141 | rma | invariantset | pmonly | mas | |
| 0.62 | 0.65 | 0.62 | 0.58 | 0.63 | 0.67 | 0.57 | 0.58 | 0.6134 | mas | quantiles | pmonly | medianpolish | |
| 0.61 | 0.65 | 0.61 | 0.57 | 0.63 | 0.67 | 0.57 | 0.59 | 0.6129 | rma | constant | pmonly | medianpolish | |
| 0.62 | 0.65 | 0.62 | 0.58 | 0.63 | 0.67 | 0.56 | 0.58 | 0.6128 | rma | quantiles | pmonly | medianpolish | rma |
| 0.62 | 0.64 | 0.61 | 0.58 | 0.63 | 0.67 | 0.56 | 0.58 | 0.6118 | rma | invariantset | pmonly | medianpolish | |
| 0.61 | 0.64 | 0.62 | 0.57 | 0.63 | 0.67 | 0.57 | 0.58 | 0.6114 | mas | invariantset | pmonly | medianpolish | |
| 0.62 | 0.64 | 0.61 | 0.57 | 0.63 | 0.66 | 0.57 | 0.58 | 0.6112 | none | quantiles | pmonly | medianpolish | |
| 0.62 | 0.64 | 0.61 | 0.57 | 0.63 | 0.66 | 0.57 | 0.58 | 0.6109 | none | invariantset | pmonly | medianpolish | |
| 0.62 | 0.64 | 0.61 | 0.56 | 0.63 | 0.66 | 0.55 | 0.59 | 0.6088 | none | constant | pmonly | medianpolish | |

Values are given for each strain individually and as an average across strains. The right part of the table describes the combination of normalization steps to yield the different methods. Boldface highlight the correlation coefficients for the commonly employed normalization methods.

The original reasoning behind using a mismatch oligo on the chip was to correct the perfect match signal intensities for unspecific binding. However, numerous studies have shown that often mismatch signal intensities exceed perfect match intensities and that mismatch oligos poorly respond to changes in the target gene expression (6). Thus, one would expect that incorporating mismatch oligos in the normalization procedure rather increase noise than specificity, which would explain

the advantage of the perfect match only method for replicate arrays.

Taken together, our results suggest that correlations among co-regulated genes and correlations between replicate experiments are sensitive to different processes during the normalization procedure. In the case of replicate arrays the most important aspect is that the expression value of one gene on array 1 is similar to the expression level of the

Table 3. Spearman correlation coefficients between replicate samples

| MG <i>r</i> | MGF <i>r</i> | DH <i>r</i> | DHF <i>r</i> | Average | Method Background | Normalize | pmcorrect | Summary | Common name |
|----------------|-----------------|----------------|-----------------|--------------|----------------------|---------------------|---------------|---------------------|----------------|
| 0.9531 | 0.9853 | 0.986 | 0.9657 | 0.973 | gcrma | invariantset | pmonly | medianpolish | |
| 0.9571 | 0.9821 | 0.9842 | 0.9634 | 0.972 | none | quantiles | pmonly | medianpolish | |
| 0.9646 | 0.98 | 0.9827 | 0.9577 | 0.971 | mas | invariantset | pmonly | medianpolish | |
| 0.9549 | 0.9822 | 0.9843 | 0.9636 | 0.971 | none | invariantset | pmonly | medianpolish | |
| 0.9628 | 0.9795 | 0.9823 | 0.9568 | 0.970 | mas | quantiles | pmonly | medianpolish | |
| 0.9622 | 0.9792 | 0.9819 | 0.9567 | 0.970 | mas | constant | pmonly | medianpolish | |
| 0.9348 | 0.9867 | 0.9869 | 0.9682 | 0.969 | gcrma | quantiles | pmonly | medianpolish | gcrma |
| 0.9547 | 0.981 | 0.982 | 0.9587 | 0.969 | none | constant | pmonly | medianpolish | |
| 0.9489 | 0.9832 | 0.9831 | 0.9607 | 0.969 | gcrma | constant | pmonly | medianpolish | |
| 0.9578 | 0.9792 | 0.9821 | 0.9568 | 0.969 | rma | invariantset | pmonly | medianpolish | |
| 0.9567 | 0.9788 | 0.9819 | 0.9561 | 0.968 | rma | constant | pmonly | medianpolish | |
| 0.9542 | 0.979 | 0.982 | 0.9571 | 0.968 | rma | quantiles | pmonly | medianpolish | rma |
| 0.9734 | 0.9696 | 0.9784 | 0.9493 | 0.968 | mas | invariantset | mas | medianpolish | |
| 0.9567 | 0.9729 | 0.9813 | 0.9565 | 0.967 | mas | quantiles | pmonly | liwong | |
| 0.9552 | 0.9728 | 0.9807 | 0.9577 | 0.967 | mas | constant | pmonly | liwong | |
| 0.9554 | 0.9722 | 0.9811 | 0.9575 | 0.967 | rma | constant | pmonly | liwong | |
| | 0.9719 | 0.9786 | 0.949 | 0.967 | mas | quantiles | mas | medianpolish | |
| 0.9557 | 0.9721 | 0.9802 | 0.9565 | 0.966 | mas | invariantset | pmonly | liwong | |
| 0.9616 | 0.9733 | 0.979 | 0.9507 | 0.966 | mas | constant | mas | medianpolish | |
| 0.9687 | 0.9705 | 0.9771 | 0.9443 | 0.965 | none | invariantset | pmonly | mas | |
| 0.9684 | 0.9705 | 0.977 | 0.9446 | 0.965 | none | quantiles | pmonly | mas | |
| 0.9411 | 0.978 | 0.9805 | 0.9588 | 0.965 | gcrma | quantiles | pmonly | liwong | |
| 0.9557 | 0.9727 | 0.9797 | 0.9502 | 0.965 | none | invariantset | mas | medianpolish | |
| 0.9519 | 0.9704 | 0.9791 | 0.9569 | 0.965 | mas | constant | mas | liwong | |
| 0.952 | 0.9749 | 0.9801 | 0.9511 | 0.965 | none | constant | mas | medianpolish | |
| 0.9545 | 0.969 | 0.9794 | 0.9549 | 0.964 | mas | quantiles | mas | liwong | |
| 0.9542 | 0.9732 | 0.9797 | 0.9505 | 0.964 | none | quantiles | mas | medianpolish | |
| 0.9453 | 0.9732 | 0.9816 | 0.957 | 0.964 | rma | quantiles | pmonly | liwong | |
| 0.9641 | 0.9703 | 0.9769 | 0.944 | 0.964 | none | constant | pmonly | mas | |
| 0.9626 | 0.9707 | 0.9769 | 0.945 | 0.964 | rma | quantiles | pmonly | mas | |
| 0.9455 | 0.9722 | 0.9808 | 0.9565 | 0.964 | rma | invariantset | pmonly | liwong | |
| 0.9459 | 0.9729 | 0.9795 | 0.9558 | 0.964 | none | quantiles | pmonly | liwong | |
| 0.9615 | 0.9706 | 0.9768 | 0.9449 | 0.963 | rma | invariantset | pmonly | mas | |
| 0.9412 | 0.977 | 0.979 | 0.956 | 0.963 | gcrma | constant | pmonly | liwong | |
| 0.955 | 0.9662 | 0.978 | 0.9538 | 0.963 | mas | invariantset | mas | liwong | |
| 0.9443 | 0.9719 | 0.9792 | 0.9556 | 0.963 | none | constant | pmonly | liwong | |
| 0.9573 | 0.9705 | 0.9769 | 0.9446 | 0.962 | rma | constant | pmonly | mas | |
| 0.937 | 0.9723 | 0.9797 | 0.9588 | 0.962 | none | constant | mas | liwong | |
| 0.9414 | 0.9751 | 0.9779 | 0.9521 | 0.962 | gcrma | invariantset | pmonly | liwong | |
| 0.9359 | 0.9713 | 0.9804 | 0.9579 | 0.961 | none | quantiles | mas | liwong | |
| 0.9351 | 0.9693 | 0.9801 | 0.9569 | 0.960 | none | invariantset | mas | liwong | |
| 0.9336 | 0.9724 | 0.9789 | 0.9547 | 0.960 | none | invariantset | pmonly | liwong | Li-Wong |
| 0.946 | 0.9708 | 0.977 | 0.9447 | 0.960 | mas | constant | pmonly | mas | |
| 0.9464 | 0.9706 | 0.9771 | 0.9445 | 0.960 | mas | invariantset | pmonly | mas | |
| 0.9458 | 0.9708 | 0.977 | 0.9447 | 0.960 | mas | quantiles | pmonly | mas | |
| 0.9773 | 0.9585 | 0.9686 | 0.931 | 0.959 | gcrma | quantiles | pmonly | mas | |
| 0.9758 | 0.9589 | 0.9685 | 0.9315 | 0.959 | gcrma | invariantset | pmonly | mas | |
| 0.971 | 0.9586 | 0.9684 | 0.9315 | 0.957 | gcrma | constant | pmonly | mas | |
| 0.9588 | 0.9591 | 0.9705 | 0.9325 | 0.955 | none | constant | mas | mas | |
| 0.9585 | 0.9573 | 0.9697 | 0.9324 | 0.954 | none | quantiles | mas | mas | |
| 0.9587 | 0.9563 | 0.9693 | 0.9317 | 0.954 | none | invariantset | mas | mas | |
| 0.9562 | 0.9564 | 0.969 | 0.9306 | 0.953 | mas | quantiles | mas | mas | |
| 0.9455 | 0.9544 | 0.9685 | 0.9297 | 0.950 | mas | invariantset | mas | mas | |
| 0.9362 | 0.9584 | 0.9697 | 0.9318 | 0.949 | mas | constant | mas | mas | mas5.0 |

Values are given for each strain individually and as an average across strains. The right part of the table describes the combination of normalization steps to yield the different methods. Boldface highlight the correlation coefficients for the commonly employed normalization methods.

same gene on array 2. Everything that contributes to the consistency of a gene's expression level across the arrays is beneficial in the normalization method, whereas the actual expression level of the gene is only of secondary importance. The opposite is true in the case where co-expression among different genes within an array is of interest. Here, algorithms that predict actual expression levels correctly are superior to those that mainly focus on consistency between chips.

Our study implies that researchers trying to identify networks of co-expressed genes within a single array are better advised to use the *Li-Wong* summary method. On the other hand, for the detection of differentially expressed genes, the *rma/gcrma* normalization method is superior. A potential limitation of the *Li-Wong* method to detect co-expression is that a relatively large number of chips have to be available for the model to yield reliable fitted values. However, our dataset consisted of 8 chips,

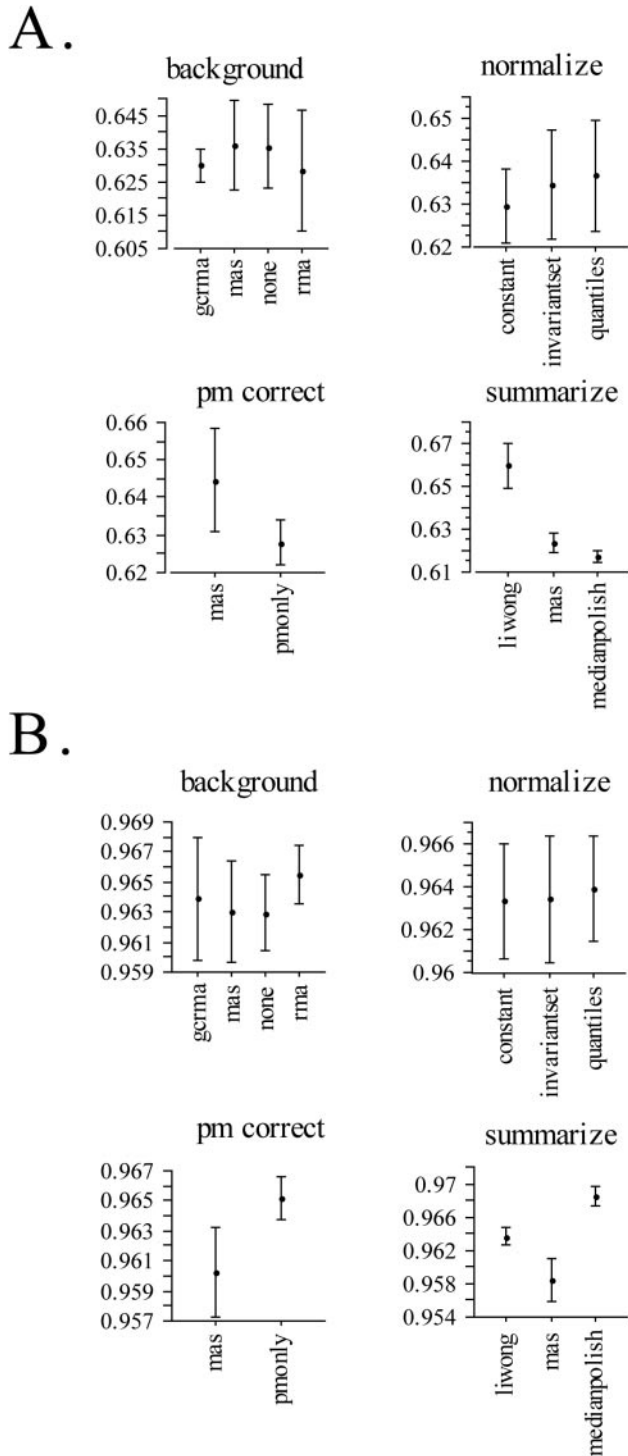


Figure 2. The influence of different methods is shown for each step in the normalization procedure independently. (A) Mean and 95% confidence intervals of Spearman rank correlation coefficients calculated between the two member genes of an operon. (B) Mean and 95% confidence intervals of Spearman rank correlation coefficients calculated between all genes on replicate arrays.

each of 4404 genes. On average, only 15 probes did not converge using the *Li-Wong* summary method, suggesting that a rather limited number of arrays are sufficiently informative.

ACKNOWLEDGEMENTS

We are grateful to Xinmin Li from the microarray facility at the University of Chicago for expert technical help. Trevor Price provided valuable statistical advice and comments on the manuscript. This work was supported by Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) grants to C.S. and an Emmy-Noether fellowship by the DFG to B.H. Funding to pay the Open Access publication charges for this article was provided by the German Science Foundation (DFG).

Conflict of interest statement. None declared.

REFERENCES

- Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680.
- Affymetrix,I. (2002) Statistical Algorithms Description Document. <http://www.affymetrix.com/support/technical/whitepapers.affx>
- Li,C. and Wong,W.H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA*, **98**, 31–36.
- Li,C. and Hung Wong,W. (2001) Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol.*, **2**, RESEARCH0032.1-11.
- Wu,Z., Irizarry,R.A., Gentleman,R., Murillo,F.M. and Spencer,F. (2004) *A Model Based Background Adjustment for Oligonucleotide Expression Arrays*. Working Papers, Department of Biostatistics, Johns Hopkins University.
- Irizarry,R.A., Hobbs,B., Collin,F., Beazer-Barclay,Y.D., Antonellis,K.J., Scherf,U. and Speed,T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
- Shedden,K., Chen,W., Kuick,R., Ghosh,D., Macdonald,J., Cho,K.R., Giordano,T.J., Gruber,S.B., Fearon,E.R., Taylor,J.M. *et al.* (2005) Comparison of seven methods for producing Affymetrix expression scores based on False Discovery Rates in disease profiling data. *BMC Bioinformatics*, **6**, 26.
- Hoffmann,R., Seidl,T. and Dugas,M. (2002) Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. *Genome Biol.*, **3**, RESEARCH0033.1-11.
- Irizarry,R.A., Bolstad,B.M., Collin,F., Cope,L.M., Hobbs,B. and Speed,T.P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, e15.
- Galfalvy,H., Erraji-Benchekroun,L., Smyrniotopoulos,P., Pavlidis,P., Ellis,S., Mann,J.J., Sibille,E. and Arango,V. (2003) Sex genes for genomic analysis in human brain: internal controls for comparison of probe level data extraction. *BMC Bioinformatics*, **4**, 37.
- Blattner,F.R., Plunkett,G.,III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.
- Bolstad,B.M. (2004) affy: Built-in Processing Methods. <http://www.bioconductor.org>.
- Salgado,H., Santos-Zavaleta,A., Gama-Castro,S., Millan-Zarate,D., Blattner,F.R. and Collado-Vides,J. (2000) RegulonDB (version 3.0): transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res.*, **28**, 65–67.
- Huerta,A.M., Glasner,J.D., Gutiérrez-Ríos,R.M., Blattner,F.R. and Collado-Vides,J. (2002) GETools: gene expression analysis for transcriptome experiments in *E.coli*. *Trends Genet.*, **18**, 217–218.
- Bammler,T., Beyer,R.P., Bhattacharya,S., Boorman,G.A., Boyles,A., Bradford,B.U., Bumgarner,R.E., Bushel,P.R., Chaturvedi,K., Choi,D. *et al.* (2005) Standardizing global gene expression analysis between laboratories and across platforms. *Nature Methods*, **2**, 351–356.
- Wang,H., He,X., Band,M., Wilson,C. and Liu,L. (2005) A study of inter-lab and inter-platform agreement of DNA microarray data. *BMC Genomics*, **6**, 71.

17. Zakharkin,S.O., Kim,K., Mehta,T., Chen,L., Barnes,S., Scheirer,K.E., Parrish,R.S., Allison,D.B. and Page,G.P. (2005) Sources of variation in Affymetrix microarray experiments. *BMC Bioinformatics*, **6**, 214.
18. Tjaden,B., Haynor,D.R., Stolyar,S., Rosenow,C. and Kolker,E. (2002) Identifying operons and untranslated regions of transcripts using *Escherichia coli* RNA expression analysis. *Bioinformatics*, **18**, S337–S344.
19. Perez-Roger,I., Garcia-Sogo,M., Navarro-Avino,J.P., Lopez-Acedo,C., Macian,F. and Armengod,M.E. (1991) Positive and negative regulatory elements in the dnaA-dnaN-recF operon of *Escherichia coli*. *Biochimie*, **73**, 329–334.